UNIVERSITY OF HAWAII
LIBRARY
Jun 14 '52

The
PHILOSOPHICAL MAGAZINE

FIRST PUBLISHED IN 1798

OL. 43 SEVENTH SERIES No. 340 May, 1952

A Journal of Theoretical Experimental and Applied Physics

EDITOR
PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

EDITORIAL BOARD
SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.
SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.
PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.

PRICE 15s. 0d.
Annual Subscription £8 0s. 0d. payable in advance

ED AND PUBLISHED BY TAYLOR & FRANCIS LTD., RED LION COURT, FLEET ST., LONDON, E.C.4.

ADVANCES IN PHYSICS

A QUARTERLY SUPPLEMENT OF THE PHILOSOPHICAL MAGAZINE

On 1st January, 1952, the first number of this new Quarterly Supplement to the Philosophical Magazine was published. The aim of this Supplement will be to give those interested in physics comprehensive and authoritative accounts of recent important developments. It is felt by the Editor that in view of the rapid advances in many branches of physics, scientists will welcome a journal devoted to articles of this type.

VOLUME 1 APRIL 1952 NUMBER 2

Theories of Helium II. By Dr. R. B. Dingle (Cambridge University).

Wave Propagation and Flow in Liquid Helium II.
By Professor K. R. Atkins (University of Toronto, Canada).

Properties of Helium Three at Low Temperatures.
By Dr. J. G. Daunt (Ohio State University, U.S.A.).

Subject to alteration up to the time of going to Press

PRICE *per part* 15/- *plus postage* PRICE *per annum* £2 15s. 0d. *post free*

Editor: PROFESSOR N. F. MOTT, M.A., D.Sc., F.R.S.

Editorial Board: SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.
PROFESSOR A. M. TYNDALL, C.B.E., D.Sc., F.R.S.
SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

Printed and Published by
TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET ST., LONDON, E.C.4

XLV. *Angular Distribution in (d, p) and (d, n) Reactions*

By A. B. Bhatia,[*] Kun Huang,[†] R. Huby and H. C. Newns
Department of Theoretical Physics, University of Liverpool[‡]

[Received February 12, 1952]

Abstract

The angular distribution of the particles emitted in (d, p) or (d, n) reactions is calculated, assuming a stripping process, by means of the Born approximation. The result resembles in most respects that of more complicated calculations by S. T. Butler, but there are certain differences which are discussed in relation to experiments which have been, or may be done. The particles are emitted mainly at fairly small angles to the incident direction, and the precise shape of the distribution is simply related to changes of angular momentum and parity occurring. The principal unknowns in the formulae are matrix elements, which may be estimated for sufficiently high energies by the statistical method.

§1. Introduction

Many measurements of the angular distribution in (d, n) and (d, p) reactions have been made recently using deuterons of energy in the region 8–15 mev (e.g. Holt and Young (1950), Burrows *et al.* (1950), Hughes (1951), El Bedewi *et al.* (1951 a, b), Gove (1951), Allen *et al.* (1951)).

The most pronounced feature in general is the existence of one or more peaks at fairly small angles to the incident direction. A simple theory to account for these results was developed by K. Huang and A. B. Bhatia, and briefly reported on by R. Huby (1950). This theory is based on the stripping mechanism as proposed by Serber (1947) to account for (d, p) and (d, n) results (including the angular distribution) at much higher energies (~200 mev deuterons). At the lower deuteron energy now under consideration a simple Born approximation is applied to the stripping model, and a generally satisfactory form of angular distribution results. For outmoving neutrons or protons of definite energy, the form of the angular distribution is related in a simple way to the change of angular momentum and parity produced in the target nucleus by the capture of a nucleon from the incident deuteron.

The use of the Born approximation for a nuclear process at the energies concerned is difficult to justify theoretically; though as the form of the angular distribution turns out to be determined largely by such general features as changes of angular momentum and parity, it may be hoped that the results are largely independent of the details of treatment, rather in the same way as the dispersion formula for nuclear reactions can be obtained by a simple second-order perturbation treatment (e.g. Bethe 1937) as well as by more correct methods.

---

[*] Now at Department of Mathematical Physics, University of Edinburgh.
[†] Now at Department of Physics, Peking University.
[‡] Communicated by the Authors.

Butler (1950, 1951) has also given a theory for the angular distribution in (d, p) and (d, n) reactions, based on the stripping mechanism. The mathematical technique used is not a perturbation one, but of a more penetrating kind involving the smooth joining up of wave functions for the captured nucleon at a radius r_0, and thus allowing rigorously for the strong interaction of this nucleon with the nucleus. However, the theory appears still to rely on some approximations whose effect is difficult to assess, notably the assumption that there is a region near the nucleus (radius$\sim r_0$) where one may neglect the interactions of the captured nucleon both with its former partner in the deuteron and with the target nucleus. It is believed that the development given in the general section (§ 3) of Butler's (1951) paper contains some errors, though his final formula (34) is correct. For these reasons, and because of the complexity of the calculations of Butler, it is thought that the much simpler calculations of the Born approximation may be of interest, demonstrating as they do very clearly what actually happens in the reaction.

The formula obtained with the Born approximation (18 below) is very similar to Butler's in its general features. It contains a radial parameter R which is more arbitrary to fit correctly than the corresponding one (r_0) in Butler's formula, but otherwise it usually gives equally good agreement with experiment.

The angular distributions obtained with the two formulae differ most when the final nucleus is left in a state only just stable, or slightly unstable, against nucleon decay. The preliminary indications are that in such a case the Born formula may give the better agreement with experiment, but further experimental evidence is needed (see § 6). It should be noted that neither formula takes account of the Coulomb repulsion between the deuteron and the nucleus, so that it is necessary that the deuteron energy be sufficiently above the Coulomb barrier. Coulomb effects will presumably persist up to higher energies in the (d, n) than in the (d, p) reaction, because in the former the proton bearing the charge has to penetrate right to the nucleus, whereas in the latter the reaction can take place without the proton's ever coming so close (cf. Oppenheimer and Phillips (1935)).

The formula derived below (18) has been applied by H. C. Newns to experiments by El Bedewi *et al.* (1951 a, b) and El Bedewi (1952).

The apparent success of the Born approximation for calculating angular distributions in the stripping reaction suggests its application also to other reactions which may depend on a related mechanism. This has already been tried for the inelastic scattering of deuterons, and results obtained consistent with experiment (Huby and Newns 1951).

The formula of Huang and Bhatia (18) is derived in § 2, discussed in § 3 and compared with that of Butler in § 4. In § 5 it is shown how the formula can be used to obtain information about the state of the final nucleus. An example is discussed in § 6. The Appendix deals with results obtainable by recourse to the statistical theory of nuclei.

§ 2. DERIVATION OF FORMULA FOR THE CROSS-SECTION

The discussions following will refer specifically to the (d, p) reaction, but the (d, n) reaction is embraced by interchanging the roles of neutron and proton throughout.

In the stripping theory of the (d, p) reaction, it is assumed that the proton in the deuteron never comes into contact with the target nucleus : the neutron in the deuteron strikes the nucleus and is absorbed, while the proton continues its initial motion undisturbed, and emerges. We therefore only take into account an interaction of the neutron with the nucleus. The Born approximation for the differential cross-section is then given by :

$$\sigma=\frac{1}{4\pi^2\hbar^4}M_p^*M_d^*\frac{k_p}{k_d}|I|^2, \qquad \ldots\ldots\ldots \quad (1)$$

where

$$I=\int\chi_f^*(\mathbf{r}_n,\sigma_n,\xi)\,\chi_p^*(\sigma_p)e^{-i\mathbf{k}_p\cdot\mathbf{r}_p'}\,V(\mathbf{r}_n,\sigma_n,\xi)\,\chi_i(\xi)\,\chi_d(\mathbf{r}_n-\mathbf{r}_p,\sigma_n,\sigma_p)$$
$$e^{[i\mathbf{k}_d\cdot(\mathbf{r}_n+\mathbf{r}_p)]/2}\,d\mathbf{r}_n\,d\mathbf{r}_p\,d\sigma_n\,d\sigma_p\,d\xi. \qquad \ldots\ldots\ldots \quad (2)$$

The symbols have the meaning :

M_p^*=reduced mass of proton$=\dfrac{M_pM_f}{M_p+M_f}$,

M_d^*=reduced mass of deuteron$=\dfrac{M_dM_i}{M_d+M_i}$,

where M_p, M_n, M_d, M_i, M_f, are the masses of proton, neutron, deuteron, initial nucleus and final nucleus respectively ; $\mathbf{k}_p=(M_p^*\mathbf{v}_p)/\hbar$, $\mathbf{k}_d=(M_d^*\mathbf{v}_d)/\hbar$, where $\mathbf{v}_p$ is the velocity of the outgoing proton relative to the centre of gravity of the final nucleus, and $\mathbf{v}_d$ is the velocity of the incoming deuteron in the laboratory system ; $\mathbf{r}_n$, $\mathbf{r}_p$ are the coordinates of the neutron and proton relative to the centre of gravity of the initial nucleus ; $\mathbf{r}_p'$ are the coordinates of the proton relative to the centre of gravity of the final nucleus $(\mathbf{r}_p'=\mathbf{r}_p-(M_n/M_f)\mathbf{r}_n)$; σ_n, σ_p are the spin coordinates of neutron and proton ; ξ represents all the internal coordinates of the initial nucleus ; χ_p, χ_d, χ_i, χ_f are internal wave functions for the proton, deuteron, initial and final nuclei respectively ; $V(\mathbf{r}_n, \sigma_n, \xi)$ is the interaction operator between the neutron and the particles in the initial nucleus.

We assume that the process of capture of the neutron takes place close to the surface of the initial nucleus, or, more precisely, in some narrow shell just outside the surface, where we can roughly imagine the attraction of the nucleus for the neutron to outweigh that of the proton in the deuteron. Let the mean radius of this shell be R, which ought to be approximately the radius of the nucleus plus the range of nuclear forces. Mathematically, this assumption implies that the major contribution to the integral (2) comes from a small range of r_n close to R, and so in the factors of the integrand which represent the incident wave, we may replace $\mathbf{r}=(r_n, \theta_n, \phi_n)$ by $\mathbf{R}=(R, \theta_n, \phi_n)$ and obtain

$$I=\int \chi_p^*(\sigma_p) e^{-i\mathbf{k}_p \cdot \mathbf{r}_p'} \chi_d(\mathbf{R}-\mathbf{r}_p, \sigma_n, \sigma_p) e^{[i\mathbf{k}_d \cdot (\mathbf{R}+\mathbf{r}_p)]/2}$$
$$\{\int \chi_f^*(\mathbf{r}_n, \sigma_n, \xi) V(\mathbf{r}_n, \sigma_n, \xi) \chi_i(\xi) r_n^2 dr_n d\xi\} d\mathbf{r}_p d\sigma_p d\sigma_n d\omega_n. \quad . \quad . \quad (3)$$

The expression in the brackets is a function of the angular coordinates and the spin of the neutron. It may thus be expanded as follows :

$$\int \chi_f^*(\mathbf{r}_n, \sigma_n, \xi) V(\mathbf{r}_n, \sigma_n, \xi) \chi_i(\xi) r_n^2 dr_n d\xi$$
$$= \sum_{l_n, m_n, \mu_n} \langle f|V|i, l_n, m_n, \mu_n\rangle Y_{l_n}^{m_n *}(\theta_n, \phi_n) \chi_{\mu_n}^*(\sigma_n), \quad . \quad . \quad . \quad (4)$$

where $Y_{ln}^{m_n}$ is a spherical harmonic, and $\chi_{\mu_n}(\sigma_n)$ is a spin eigenfunction. The expansion coefficients are given by :

$$\langle f|V|i, l_n, m_n, \mu_n\rangle=\int \chi_f^*(\mathbf{r}_n, \sigma_n, \xi) V(\mathbf{r}_n, \sigma_n, \xi) \chi_i(\xi) Y_{l_n}^{m_n}$$
$$(\theta_n, \phi_n) \chi_{\mu_n}(\sigma_n) d\mathbf{r}_n d\sigma_n d\xi. \quad . \quad . \quad . \quad (5)$$

Substitution of (4) in (3) yields

$$I= \sum_{l_n, m_n, \mu_n} \langle f|V|i, l_n, m_n, \mu_n\rangle \langle l_n, m_n, \mu_n, p|d\rangle, \quad . \quad . \quad . \quad (6)$$

where

$$\langle l_n, m_n, \mu_n, p|d\rangle=\int Y_{l_n}^{m_n *}(\theta_n, \phi_n) \chi_{\mu_n}^*(\sigma_n) \chi_p^*(\sigma_p) e^{-i\mathbf{k}_p \cdot \mathbf{r}_p'}$$
$$\times \chi_d(\mathbf{R}-\mathbf{r}_p, \sigma_n, \sigma_p) e^{[i\mathbf{k}_d \cdot (\mathbf{R}+\mathbf{r}_p)]/2} d\mathbf{r}_p d\sigma_p d\sigma_n d\omega_n. \quad . \quad (7)$$

Now the components of angular momentum of the various internal wave functions viz. $\mu_p, \mu_d, \mu_i, \mu_f$ are experimentally unspecified, and so the quantity $|I|^2$ required in (1) is rather an average over the initial and sum over the final components, i.e., with the aid of (6) we modify $|I|^2$ to :

$$|I|^2=\frac{1}{3(2j_i+1)} \sum_{\substack{\mu_p, \mu_d \\ \mu_i, \mu_f}} \left| \sum_{l_n, m_n, \mu_n} \langle \mu_f|V|\mu_i, l_n, m_n, \mu_n\rangle \langle l_n, m_n, \mu_n, \mu_p|\mu_d\rangle \right|^2. \quad . \quad . \quad . \quad (8)$$

Here the insertion of symbols such as μ_p in the Dirac brackets $\langle | \rangle$ indicates that internal wave functions χ_{μ_p} etc. with the specified angular momentum are to be used in the integrals (5) and (7).

(8) can be simplified if the d-state component of the deuteron's wave function is neglected, i.e. if χ_d is a product of a triplet spin factor and an s-state orbital factor :*

$$\chi_{\mu_d}(\mathbf{R}-\mathbf{r}_p, \sigma_n, \sigma_p)=\psi(|\mathbf{R}-\mathbf{r}_p|) S_{\mu_d}(\sigma_n, \sigma_p).$$

---

* If the d admixture were allowed for, the result would probably be more complicated, and the selection rules governing l_n (§ 3) would probably be less stringent. These effects are, however, expected to be of small magnitude.

Then (7) becomes

$$\langle l_n, m_n, \mu_n, \mu_p \,|\, \mu_d \rangle = g_{l_n}{}^{m_n} \langle \mu_n, \mu_p \,|\, \mu_d \rangle, \quad . \; . \; . \; . \quad (9)$$

where

$$g_{l_n}{}^{m_n} = \int Y_{l_n}{}^{m_n *}(\theta_n, \phi_n) \mathrm{e}^{-i\mathbf{k}_p \cdot \mathbf{r}_p} \psi(|\mathbf{R} - \mathbf{r}_p|) \mathrm{e}^{[i\mathbf{k}_d(\mathbf{R}+\mathbf{r}_p)]/2} d\omega_n d\mathbf{r}_p, \quad . \quad (10)$$

and

$$\langle \mu_p, \mu_n \,|\, \mu_d \rangle = \int \overset{*}{\chi}_{\mu_p}(\sigma_p) \overset{*}{\chi}_{\mu_n}(\sigma_n) S_{\mu_d}(\sigma_n, \sigma_p)\, d\sigma_n\, d\sigma_p. \quad . \; . \quad (11)$$

After substituting (9) in (8), we can easily perform the summation over μ_p and μ_d with the result :

$$|I|^2 = \frac{1}{2(2j_i+1)} \sum_{\mu_n, \mu_i, \mu_f} \Big| \sum_{l_n, m_n} \langle \mu_f | V | \mu_i, l_n, m_n, \mu_n \rangle g_{l_n}{}^{m_n} \Big|^2. \quad . \quad (12)$$

The summation over μ_n, μ_i and μ_f can now be carried out with the aid of the group-theoretical properties of the wave functions concerned in the definition of the Dirac bracket expressions. The result is

$$|I|^2 = \frac{1}{2(2j_i+1)} \sum_{l_n, m_n} \frac{\Lambda_{l_n}}{(2l_n+1)} |g_{l_n}{}^{m_n}|^2, \quad . \; . \; . \; . \quad (13)$$

where

$$\Lambda_{l_n} = \sum_{\substack{m_n', \mu_n \\ \mu_i, \mu_f}} \Big| \langle \mu_f | V | \mu_i, l_n, m_n', \mu_n \rangle \Big|^2. \quad . \; . \; . \; . \quad (14)$$

$g_{l_n}{}^{m_n}$ can readily be evaluated from (10). If the principal axis for determining l_n is taken in the direction of

$$\mathbf{k} = \mathbf{k}_d - \frac{M_i}{M_f} \mathbf{k}_p, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (15)$$

(note $k = \left[\left(k_d - \frac{M_i}{M_f} k_p \right)^2 + 4 \frac{M_i}{M_f} k_p k_d \sin^2 \tfrac{1}{2}\theta \right]^{1/2}$, where θ is the angle between $\mathbf{k}_p$ and $\mathbf{k}_d$), then we find :

$$g_{l_n}{}^{m_n} = \delta_{m_n, 0} [4\pi(2l_n+1)]^{1/2} i^{l_n} G(|\mathbf{k}_p - \tfrac{1}{2}\mathbf{k}_d|) \left(\frac{\pi}{2kR} \right)^{1/2} J_{l_n+\frac{1}{2}}(kR), \quad . \; . \quad (16)$$

where

$$G(K) = \frac{4\pi}{K} \int_0^\infty \psi(r) r \sin Kr \,.\, dr. \quad . \; . \; . \; . \; . \; . \quad (17)$$

Substitution of (16) into (13) and then into (1) yields

$$\sigma = \frac{M_p^* M_d^*}{2\pi(2j_i+1)\hbar^4} \frac{k_p}{k_d} G^2(|\mathbf{k}_p - \tfrac{1}{2}\mathbf{k}_d|) \sum_{l_n} \Lambda_{l_n} \left[\left(\frac{\pi}{2kR} \right)^{1/2} J_{l_n+\frac{1}{2}}(kR) \right]^2. \quad . \quad (18)$$

A formula for G is obtained if we use for the deuteron spatial wave function the approximate form :

$$\psi(r) = \left(\frac{\alpha}{2\pi} \right)^{1/2} \frac{\mathrm{e}^{-\alpha r}}{r}, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (19)$$

where $\alpha = 1/\hbar \times (M_p \times \text{deuteron binding energy})^{1/2}$

$$= 0{\cdot}23 \times 10^{13}\ \text{cm}^{-1}. \quad . \; . \; . \; . \; . \; . \; . \; . \quad (20)$$

Then

$$G(|\mathbf{k}_p - \tfrac{1}{2}\mathbf{k}_d|) = \frac{2(2\pi\alpha)^{1/2}}{\alpha^2 + |\mathbf{k}_p - \tfrac{1}{2}\mathbf{k}_d|^2} = \frac{2(2\pi\alpha)^{1/2}}{\alpha^2 + (k_p - \tfrac{1}{2}k_d)^2 + 2k_p k_d \sin^2 \tfrac{1}{2}\theta}. \quad . \quad (21)$$

§ 3. Interpretation of Results

The cross-section σ (18) contains a sum of terms over a parameter l_n which may be interpreted, in view of definitions (5) and (7), as the orbital angular momentum carried by the absorbed neutron into the nucleus. There are three important factors in the formula (18) :

(i) *Deuteron Factor* G^2 (17 and 21)

The proton, which initially has the mean momentum $\hbar\mathbf{k}_d/2$ in the centre of mass frame, emerges with the momentum $\hbar\mathbf{k}_p$. The difference $\hbar(\mathbf{k}_p-\mathbf{k}_d/2)$ represents the momentum of the proton in its motion inside the deuteron, at the instant of its release due to the removal of the neutron ; and the factor $G^2(|\mathbf{k}_p-\mathbf{k}_d/2|)$ is proportional to the probability of this internal deuteron momentum. Equation (21) shows that this factor, as function of the angle θ between $\mathbf{k}_p$ and $\mathbf{k}_d$, has a forward maximum. The reason is that the larger the angle the larger the momentum difference which must be provided by the internal motion of the deuteron, and hence the smaller the probability. The factor G^2 applies equally to all transitions with different l_n.

(ii) *Neutron Centrifugal Factor* $[(\pi/2kR)^{1/2}J_{l_n+\frac{1}{2}}(kR)]^2$

From the last process, the neutron can be considered as issuing virtually from the deuteron with momentum $\hbar(\mathbf{k}_d-\mathbf{k}_p)$ in the centre of mass frame. After its capture by the target nucleus, the neutron has the mean momentum $-(M_n/M_f)\hbar\mathbf{k}_p$. The mean momentum interchanged between neutron and initial nucleus is thus $\hbar\mathbf{k}=\hbar[\mathbf{k}_d-(M_i/M_f)\mathbf{k}_p]$. The factor $[(\pi/2kR)^{1/2}J_{l_n+\frac{1}{2}}(kR)]^2$ is proportional to the mean probability that the neutron, travelling with this momentum in the frame in which the centre of gravity of the final nucleus is at rest, be found at the surface R of the nucleus, in a state in which the neutron's total orbital angular momentum is l_n.

With the aid of (15), we see that the present spherical Bessel function factor is an oscillatory function of the angle θ, the oscillations decreasing as θ increases. Provided $|k_d-(M_i/M_f)k_p|R$ is not too large, the factor has a maximum at the origin for $l_n=0$, and for all other l_n a minimum at the origin, followed by a primary maximum which decreases in magnitude and moves to larger angles as l_n increases. The reason is that the larger the angular momentum l_n, the greater is the momentum $\hbar k$ (and hence the angle θ) which is necessary for the neutron to penetrate to the radius R.

(iii) *Nuclear Matrix Element* Λ_{l_n} (5 and 14)

This is proportional to the probability of capture of a neutron which is at the surface of the nucleus with initial orbital angular momentum l_n. We may note that Λ_{l_n} is determined only by the properties of the initial and final nuclear levels, and by l_n. When j_i and j_f and the initial and final parities are given, (14) shows clearly that there will be selection rules for the values of l_n which yield non-zero values of Λ_{l_n}. It follows from conventional group theoretical arguments that the selection rules are :

(*a*) The final angular momentum j_f must be obtainable by vector addition of the initial j_i, the orbital angular momentum l_n of the absorbed neutron, and the latter's spin angular momentum 1/2.

(*b*) l_n must be even or odd, according as the parities of the initial and final levels are the same or different. These selection rules severely limit the number of terms appearing in (18), and can often bring it down to only one.

It is not easy to make predictions about the magnitudes, absolute or relative, of the Λ_{l_n}'s. Generally, it will be expected that different final states of similar constitution will have Λ_{l_n}'s of roughly the same order of magnitude. For example, for any final state in which the captured neutron is approximately moving in an orbit outside the initial nucleus in its original state, there will be only one value of l_n contributing appreciably, equal to the neutron's orbital angular momentum, even though more l_n's might be permitted by the selection rules (*a*) and (*b*). The Λ_{l_n}'s for such final states might be considerably larger than those for most others ; and the relative values of the Λ_{l_n}'s for different final states of this type would be largely governed by a statistical factor $(2j_f+1)$.

A rough estimate of the Λ_{l_n}'s may be made by the statistical method, as used by Bethe (1938) for calculations of the cross-section integrated over all angles. It is shown in the appendix that, for highly excited final states unstable against neutron emission, this yields

$$\Lambda_{l_n}=\frac{\hbar^2 k_n R^2\ (2j_f+1)\mu}{4\pi M_n^*\ \rho_{jf}(E_f)}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Here k_n is the wave number of the neutron emitted if the final nucleus decays to the initial, i.e.

$$\frac{\hbar^2 k_n{}^2}{2M_n^*}=-(Q \text{ of } (d, p) \text{ reaction leading to final state } f + \text{binding energy of deuteron}) \quad . \quad (23)$$

$\rho_{jf}(E_f)$ is the density of levels of the final nucleus in the neighbourhood of the actual final level ; counting only levels of the same angular momentum j_f and parity as the level under consideration.

μ is 1 or 2, according as only one, or both, of the vectors of lengths $(j_i+\frac{1}{2})$ and $(j_i-\frac{1}{2})$ can be added vectorially to l_n to give a vector of length j_f (otherwise, of course, $\mu=0$).

The derivation of (22) only applies for $k_nR>l_n$, but we may perhaps cautiously extrapolate it to smaller k_n, and even to imaginary k_n (corresponding to a state of the final nucleus stable against neutron emission), if we replace a factor k_nR in the numerator by unity—for an order of magnitude result.

It is also shown in the appendix how the use of (22) in the cross-section (18), for very high deuteron energy, reproduces essentially some of the results on angular distribution and energy spectrum obtained by Serber (1947), who did not analyse in detail the changes occurring in the target nucleus. For example, summation of the differential cross-section over many neighbouring final levels yields a result in which the combined effect of the centrifugal factors (spherical Bessel functions) is independent of angle, and the residual angular distribution is due to the deuteron factor alone.

§ 4. COMPARISON WITH THE FORMULA OF BUTLER (1951)

For greatest ease of comparison of formula (18) with the formula (34) of Butler (1951), the latter may be rearranged somewhat, and the notation modified, to yield :

$$\sigma = \frac{M_p^* M_d^*}{2\pi(2j_i+1)} \frac{k_p}{\hbar^4 k_d} G'^2(|\mathbf{k}_p - \tfrac{1}{2}\mathbf{k}_d|)$$

$$\times \sum_{l_n} N_{l_n} e^{\kappa_n r_0} \left[\kappa_n r_0 K_{l_n - \frac{1}{2}}(\kappa_n r_0) \left(\frac{\pi}{2kr_0}\right)^{1/2} . J_{l_n+\frac{1}{2}}(kr_0)\right.$$

$$\left. + K_{l_n+\frac{1}{2}}(\kappa_n r_0) \left(\frac{\pi k r_0}{2}\right)^{1/2} J_{l_n-\frac{1}{2}}(kr_0) \right]^2 . \qquad (24)$$

Here G' is a function (deuteron factor) similar to G in (21), but differing a little because Butler used a more accurate form of wave function for the deuteron ground state than (19). r_0 is a nuclear radius. κ_n is defined by

$$\kappa_n = -ik_n, \qquad (25)$$

(cf. eqn. (23)), and is thus real if the final nucleus is stable against neutron decay. The case that the final nucleus is formed in a state unstable against neutron decay, and hence κ is imaginary, does not have to be treated separately in the Born calculations but it must be in Butler's. While Butler does not discuss this case fully, it is understood that the cross-section is still given by (24) provided the actual imaginary value of κ_n is inserted, and the square bracket is replaced by a modulus.

Remembering that the only variables containing the angle in (24) are k (eqn. 15) and $|\mathbf{k}_p - \frac{1}{2}\mathbf{k}_d|$, we see that the only essential difference in the angular distribution between (24) and (18) lies in adding to the spherical Bessel function $(\pi/2kr_0)^{1/2} J_{l_n+\frac{1}{2}}(kr_0)$ another term proportional to $(\pi k r_0/2)^{1/2} J_{l_n-\frac{1}{2}}(kr_0)$. The effect of this is usually rather small, and either formula can usually be made to fit an experimental curve equally well, if suitable values of r_0 and R are chosen, the latter always having to be rather larger than the former.

The difference depends mainly on κ_n. It is very small for $|\kappa_n|$ large (state of final nucleus very stable or very unstable), and greatest for $|\kappa_n|$ small (state of final nucleus only just stable or just unstable), because in this case the second term in the square bracket in (24) takes control. The effects ensuing will be illustrated in an example in § 6. The conclusion is that Butler's formula behaves in a ' special ' or ' singular ' manner when $\kappa_n = 0$, while the Born formula does not, and the precise behaviour of the former will depend rather sensitively on the value of κ_n when small. It does not seem clear whether this singular behaviour is to be expected physically, and it would thus be of interest to make careful comparisons with experiment in this region.

§5. APPLICATION TO DETERMINATION OF STATES

It is to be expected that the formula (18) will only agree with experimental curves in the region of small angles where the cross-section is large ; since the constructive interference of many partial waves of definite orbital angular momentum for the proton and deuteron, which is really responsible for the maximum cross-section, should be more reliable than the details of the destructive interference which makes the cross-section small at large angles ; and, in addition, there will probably be a background of neutrons produced by the competing process involving formation and decay of a compound nucleus, and these will tend to mask the stripping cross-section where it is small.*

To compare (18) with an experimental curve, it is necessary to know which Λ_{l_n}'s are present, and their value. For certain initial and final states more than one l_n may be permitted by the selection rules of § 3 (iii)—though the values permitted must be either all even or all odd. However, in practice, it has been found possible in most cases to obtain a satisfactory fit with one l_n only. When this has been found, if the parity and spin of the initial state are known then the selection rules fix the parity of the final level uniquely, and its spin j_f to within the limits permitted by vector addition of j_i plus l_n plus $\frac{1}{2}$.

To find l_n, the experimental curve should first be divided by the theoretical deuteron factor G^2 (see 21), so that the result is to be proportional simply to the sum in (18). A first indication of l_n can be obtained by inspection of this curve at small angles : a forward maximum will normally indicate that $l_n = 0$ is present, and a forward minimum that it is not (though these rules are not absolute). We can calculate the several spherical Bessel function terms in (18) and find which gives the best fit with experiment, if we take a suitable value of R—and this last point, to choose the right R, is a major consideration. Butler (1951) found that with his formula (24) he usually obtained good and consistent agreement using for his radius r_0 an accepted value for the radius of the nucleus plus range of nuclear forces e.g. as given in Gamow and Critchfield (1949, p. 11) :

---

* Cohen and Falk (1951) put forward experimental evidence from the energy spectrum in (*d*, *n*) reactions that the forward scattering is mainly due to stripping and that at 90° to compound nucleus. It is usually assumed that the angular distribution of particles from the compound nucleus will be fairly isotropic, but Wolfenstein's theory (1951) predicts peaking in the forward and backward directions, with symmetry about 90°. Indeed, it seems that Wolfenstein's theory could lead in extreme cases to a forward peak of sharpness comparable to that due on the stripping theory to the ' centrifugal ' factor § 3 (ii). However, it should be possible to tell whether such a forward peak is due to the stripping or the compound nucleus process, because the latter should be accompanied by a backward peak, and the former not.

$$\text{radius} = (1{\cdot}7 + 1{\cdot}22 \times A^{1/3}) \times 10^{-13}\ \text{cm}\ ; \quad . \quad . \quad . \quad . \quad (26)$$

and on the average for light nuclei he took $5{\cdot}0 \times 10^{-13}$ cm. However, to obtain good and consistent results with the Born formula, a larger value of R is found to be needed, on the average $2{\cdot}0 \times 10^{-13}$ cm greater than the Gamow value (26).

The procedure is thus to find what value of l_n, together with a value of R not very different from that just mentioned, will give the best agreement with the experimental curve. The answer has in practice almost always turned out to be unique, for a given curve.

The large value of R required is hard to justify physically, and is in contrast to the reasonable value of the r_0 required in Butler's formula. Some investigation of the role of R or r_0 has therefore been made.

One hopes that the dependence of the angular distribution on R or r_0 will be insensitive, so that their exact choice is not very important. To obtain a measure of the sensitivity, we follow up a suggestion of Professor R. E. Peierls. Take the cross-section for $l_n=0$ according to (18) and let k_0 be the value of k at which its first zero occurs. Then we may use as the gauge of sensitivity the quantity $(R/k_0)(dk_0/dR)$. Since, in fact, $k_0 \propto R$, this is just 1. The corresponding quantity $(r_0/k_0)(dk_0/dr_0)$ for Butler's formula (24) lies always between 0·78 and 1. This shows that the sensitivities of the two formulae are nearly the same.

Next we may analyse the statistics of a number of reactions for which both (18) and (24) have been fitted to the experimental curves. In each case the best values of R and r_0 to give a fit have been found, and compared to the Gamow radius (26). The results for 10 curves are that :

$$\left.\begin{array}{l}\text{for Born formula (18), } R = \text{Gamow radius} + 2{\cdot}0 \times 10^{-13} \pm 0{\cdot}50, \\ \text{for Butler's formula (24), } r_0 = \text{Gamow radius} + 0{\cdot}36 \times 10^{-13} \pm 0{\cdot}44.\end{array}\right\} . \quad (27)$$

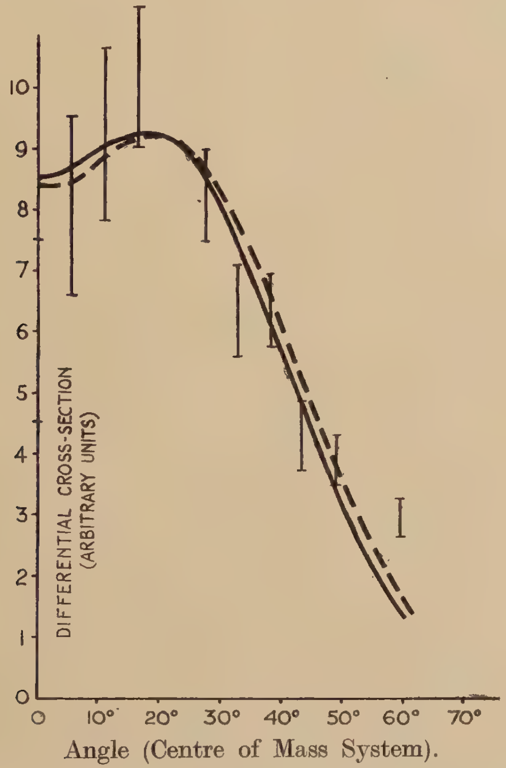
(The limits are ' probable errors ' $= 0{\cdot}67 \times$ R.M.S. deviation from mean.) Thus the spread of the best values of the radius is about the same for both the Born and Butler's theories ; but on the average Butler's radius comes out closer to the physically reasonable Gamow radius. Both methods appear in practice to be about equally reliable for finding the characteristics of the final state.

§ 6. EXAMPLE

We shall discuss only the results of the reactions $O^{16}(d, n)F^{17}$ and $O^{16}(d, n)F^{17*}$, which have been obtained by El Bedewi *et al.* (1951 a) using 8 MEV deuterons. The Q's of the respective reactions are $-1{\cdot}36$ MEV and $-1{\cdot}89$ MeV, yielding, in accordance with (25) and (23) the respective values for κ_p† of $0{\cdot}196\times10^{13}$ cm^{-1} and $0{\cdot}119\times10^{13}$ cm^{-1}, the latter in particular being unusually small.

Considering first the reaction forming F^{17} in the ground state, we find that both eqns. (18) and (24) require l_p†$=2$ for a good fit. The radius R giving the best fit by Born approximation is $6{\cdot}25\times10^{-13}$ cm (note that the Gamow radius from (26) is $4{\cdot}8\times10^{-13}$ cm, and compare with (27)); while with Butler's formula, a good result is obtained with the radius $5{\cdot}0\times10^{-13}$ cm which he recommends on the average. The results are shown in fig. 1. The insensitivity of the results to the value of the radius is shown by the curves of fig. 2, which are drawn for radii r_0 and R each increased by 1×10^{-13} cm from the values mentioned above: the fit is still quite good. The reliability of the value of l_p is indicated by the fact that, taking the Born approximation, for instance, to obtain

Fig. 1



Reaction O^{16} (d, n) F^{17} with 8 MeV deuterons. $Q=-1{\cdot}36$ MeV.
Points are measured by El Bedewi *et al.*, 1951 a.
Full curve: theory according to Born approxn., with $R=6{\cdot}25\times10^{-13}$ cm.
Broken curve: theory according to Butler, with $r_0=5\times10^{-13}$ cm.

---

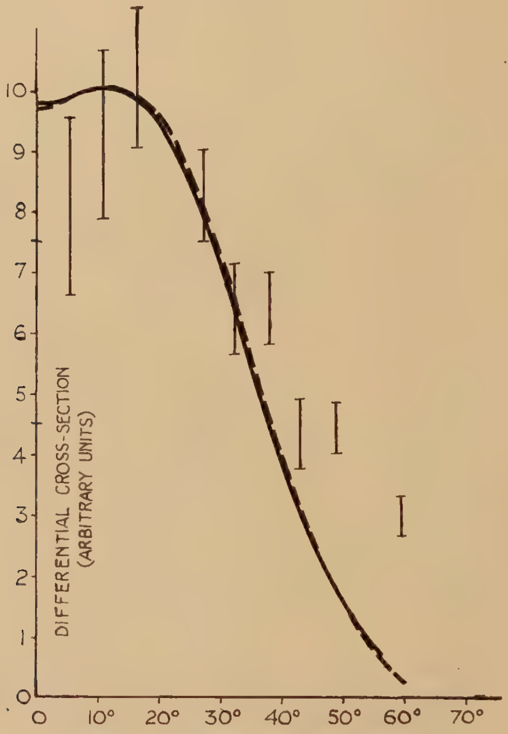
† Note that, as this is a (d, n) reaction, the roles of neutron and proton are interchanged from those in the preceding discussions.

a good fit assuming $l_p=1$ or 3, we should require respective values of R of $3{\cdot}9\times10^{-13}$ and $8{\cdot}4\times10^{-13}$ cm, neither of which conforms at all well to (27).

The conclusion to be drawn from the selection rules of § 3 (iii) is that, since the ground state of O^{16} has $j_i=0$, even parity, and since $l_p=2$, then the ground state of F^{17} has the spin $j_f=3/2$ or 5/2 and even parity.

For the first excited state, both formulae (18) and (24) require $l_p=0$. The best values of R and r_0 for a fit are respectively $7{\cdot}25\times10^{-13}$ and $4{\cdot}2\times10^{-13}$ cm, and the curves for these are shown in fig. 3. Increasing r_0 to the 'average' value quoted by Butler, $5{\cdot}0\times10^{-13}$ cm, gives poor agreement, as shown in fig. 4, whereas an increase of R in the formula (18) to $8{\cdot}05\times10^{-13}$ does not cause so much deterioration. The reason is the 'singular' behaviour of Butler's formula for small κ_p (see § 4), two of the chief features of which can be seen in this example, viz. (i) a reduction of the value of kr_0 at which the first zero of the cross-section

Fig. 2



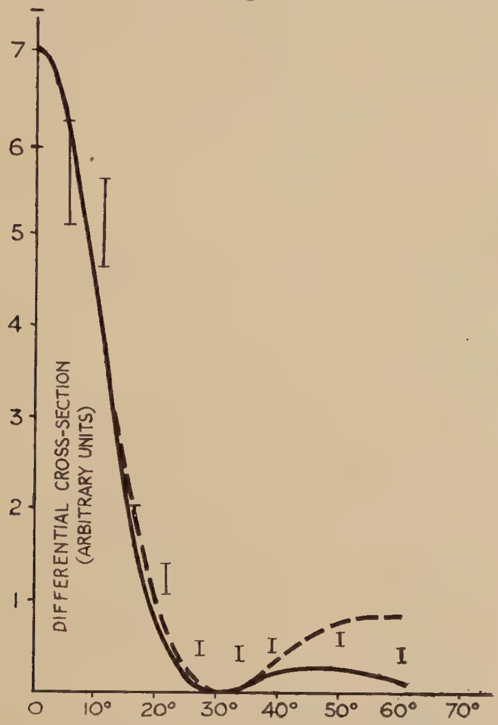
Angle (Centre of Mass System).

Reaction $O^{16}(d, n)F^{17}$ with 8 MEV deuterons. $Q=-1{\cdot}36$ MEV.
Points are measured by El Bedewi *et al.*, 1951 a.
Full curve : theory according to Born approxn., with $R=7{\cdot}25\times10^{-13}$ cm.
Broken curve : theory according to Butler, with $r_0=6\times10^{-13}$ cm.

occurs (tending to require unusually small values of r_0) and (ii) unusually large secondary maxima at large angles.*

The result $l_p=0$, together with the known spin and parity of O^{16}, leads to the result $j_f=\frac{1}{2}$, and even parity, for the first excited state of F^{17}.

Fig. 3



Angle (Centre of Mass System).

Reaction O^{16} (*d*, *n*) F^{17*} with 8 MeV deuterons. $Q=-1\cdot89$ MeV.
Points are measured by El Bedewi *et al.*, 1951 a.
Full curve : theory according to Born approxn., with $R=7\cdot25\times10^{-13}$ cm.
Broken curve : theory according to Butler, with $r_0=4\cdot2\times10^{-13}$ cm.

APPENDIX

Use of the Statistical Theory of Nuclei

(Following Bethe, 1938)

We shall express the matrix element A_{l_n} (14) in terms of the partial width for decay of the final nucleus with neutron emission, leaving the initial nucleus in its initial state (assuming this to be energetically possible).

---

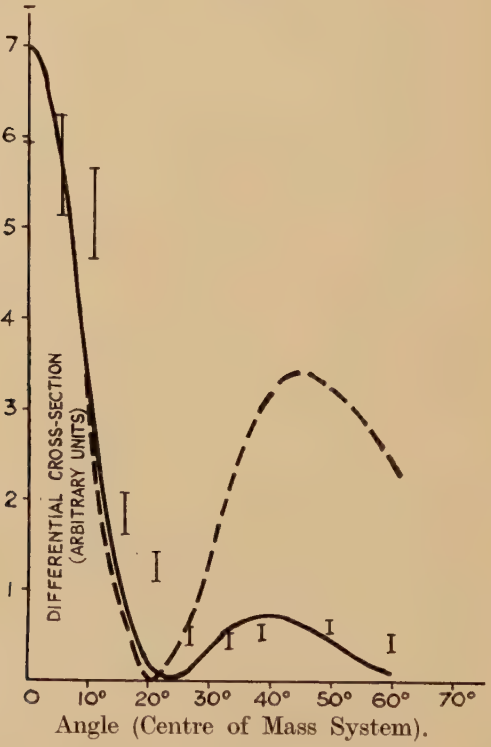
[*] Although there is as yet not much evidence under these 'singular' conditions, it is of some interest that in one other case, C^{12} (*d*, *n*) N^{13*}, (see El Bedewi *et al.* 1951 b) forming the final nucleus in its first excited state, which is slightly unstable against neutron decay ($\kappa_n=-i\times0\cdot14\times10^{13}$ cm^{-1}), it does not prove possible to obtain a really good fit for Butler's formula with any value of r_0: the nearest fit is got with $r_0=3\cdot7\times10^{-13}$ cm, whereas $r_0=5\times10^{-13}$ cm is very poor.

The partial width can then in turn be estimated in terms of a sticking probability.

The formula for the partial width for the decay process in which the neutron is emitted with angular momentum l_n is, by perturbation theory :

$$\Gamma_{j_i, l_n; j_f} = \frac{4k_n}{\hbar^2(2j_f+1)} \sum_{\substack{\mu_i, \mu_f \\ \mu_n, m_n}} \left| \int \chi^*_{\mu_i}(\xi) \chi^*_{\mu_n}(\sigma_n) Y_{l_n}{}^{m_n *}(\theta_n, \phi_n)(\pi/2k_n r_n)^{1/2} J_{l_n+\frac{1}{2}}(k_n r_n) \right.$$
$$\left. \times V(\mathbf{r}_n, \sigma_n, \xi) \chi_{\mu_f}(\mathbf{r}_n, \sigma_n, \xi) d\mathbf{r}_n \, d\sigma_n \, d\xi \right|^2 . \quad . \quad . \quad . \quad . \quad (28)$$

Fig. 4



Reaction O^{16} (d, n) F^{17*} with 8 MEV. deuterons. $Q = -1{\cdot}89$ MEV.
Points are measured by El Bedewi *et al.*, 1951 a.
Full curve : theory according to Born approxn., with $R = 8{\cdot}05 \times 10^{-13}$ cm.
Broken curve : theory according to Butler, with $r_0 = 5 \times 10^{-13}$ cm.

(The notations are those of § 2 ; k_n was defined in (23).) If we make the assumption, as in § 2, that the interaction is effective only at the radius R, so that r_n in the spherical Bessel function (28) may be replaced by R, then with the help of (5) and (14) we obtain at once :

$$\Gamma_{j_i, l_n; j_f} = \frac{4k_n A_{l_n}}{\hbar^2(2j_f+1)} \left[\left(\frac{\pi}{2k_n R} \right)^{1/2} J_{l_n+\frac{1}{2}}(k_n R) \right]^2 : \quad . \quad . \quad (29)$$

If k_nR is appreciably larger than $(l_n+\frac{1}{2})$, the asymptotic form of the Bessel function may be used, yielding on an average over a range of k_n:

$$\Gamma_{j_i, l_n ; j_f}=\frac{2\Lambda_{l_n}}{\hbar^2(2j_f+1)k_nR^2} . \qquad (30)$$

Now by the principle of detailed balance the partial width is expressible in terms of the capture probability for a neutron of wave number k_n colliding with the initial nucleus (see e.g. L. Wolfenstein, 1951)

$$\Gamma_{j_i, l_n ; j_f}=\frac{(2l_n+1)(2j_i+1)\xi_{j_f ; j_i, l_n}}{\pi(2j_f+1)\rho_{j_f}(E_f)} . \qquad (31)$$

$\rho_{j_f}(E_f)$ was defined below (23). $\xi_{j_f ; j_i, l_n}$ is the sticking probability for the neutron, if its orbital angular momentum is l_n, to combine with the initial nucleus j_i to form a compound, nucleus with the angular momentum j_f and proper parity.*

Now if we assume that the sticking probability does not depend explicitly on the final angular momentum j_f, but is determined only by statistical factors, we may write

$$\xi_{j_f ; j_i, l_n}=\frac{(2j_f+1)\mu\xi_{j_i, l_n}}{2(2l_n+1)(2j_i+1)} , \qquad (32)$$

where μ was defined below (23), and ξ_{j_i, l_n} is the sticking probability irrespective of final state, which to the present accuracy may be taken as 1. Substituting (32) in (31) and thence in (30), we obtain the final formula (22) for Λ_{l_n}.

We now investigate the cross-section obtained in measuring many proton groups together. We find first the differential cross-section $d\sigma(\theta)$ for all protons in the energy range dE_p scattered at an angle θ, from (18):

$$d\sigma(\theta)=\frac{M_p^*M_d^*}{2\pi(2j_i+1)\hbar^4}\cdot\frac{k_p}{k_d}G^2(|\mathbf{k}_p-\tfrac{1}{2}\mathbf{k}_d|)\, . \, dE_p \sum_{l_n, j_f}\Lambda_{l_n}\left[\left(\frac{\pi}{2kR}\right)^{1/2}J_{l_n+\frac{1}{2}}(kR)\right]^2 \times\rho_{j_f}(E_f)$$

(substituting from (22) and summing over j_f)

$$=\frac{G^2(|\mathbf{k}_p-\frac{1}{2}\mathbf{k}_d|)R^2}{4\pi^2\hbar^2}\frac{M_dk_pk_n}{k_d}dE_p\sum_{l_n}(2l_n+1)\left[\left(\frac{\pi}{2kR}\right)^{1/2}J_{l_n+\frac{1}{2}}(kR)\right]^2 .$$

$$=\frac{G^2(|\mathbf{k}_p-\frac{1}{2}\mathbf{k}_d|)R^2}{4\pi^2}\frac{M_dk_pk_n}{k_d}dE_p, \qquad (33)$$

since the sum over l_n is just unity. (In the above we have put $M_n^*=M_n$ etc.) The only angular dependence here is through the deuteron factor

---

* The sticking probability as here used has to contain also any inhibiting effects of barrier penetration.

G^2. (21) shows that, for high deuteron energy, this has a sharp peak, as function of $\mathbf{k}_p$, when $\mathbf{k}_p=\frac{1}{2}\mathbf{k}_d$. For important cases, then, we may put in (33)

$$k_p=k_n=\frac{k_d}{2}, \text{ and obtain}$$

$$d\sigma(\theta)=G^2(|\mathbf{k}_p-\tfrac{1}{2}\mathbf{k}_d|)\frac{R^2M_nk_d}{8\pi^2\hbar^2}\,dE_p. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

We can integrate (34) over all proton energies E_p, with the aid of (21) ; to obtain a total angular distribution $\sigma_{\text{tot}}(\theta)$ (after approximating for large k_d) :

$$\sigma_{\text{tot}}(\theta)=\frac{R^2k_d{}^2\alpha}{(\alpha^2+\frac{1}{4}k_d{}^2\sin^2\theta)^{3/2}}. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

The energy spectrum $d\sigma$ for all protons in the range dE_p is obtained similarly by integrating (33) over all angles θ with the aid of (21) :

$$d\sigma=\frac{16R^2M_p\alpha k_pk_ndE_p}{\hbar^2k_d[\alpha^2+(k_p-\frac{1}{2}k_d)^2](k_p+\frac{1}{2}k_d)^2}. \qquad . \quad . \quad . \quad . \quad . \quad (36)$$

The results (35) and (36) agree essentially with the eqns. (11) and (15) of Serber (1947) for the angular distribution and spectrum assuming a 'transparent' nucleus, i.e. neglecting entirely the interaction of the proton with the target nucleus (only in Serber's paper it would be the neutron, not proton, since he discusses (d, n) reactions).

The total cross-section σ_{tot} for all angles and all proton energies is obtained by integrating either (35) over angles or (36) over E_p :

$$\sigma_{\text{tot}}=2\pi R^2, \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

which is of the right order of magnitude, though too large.

REFERENCES

ALLEN, A. J., NECHAJ, J. F., SUN, K. H., and JENNINGS, B., 1951, *Phys. Rev.*, **81,** 536.

BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9,** 101; 1938, *Phys. Rev.*, **54,** 39.

BURROWS, H. B., GIBSON, W. M., and ROTBLAT, J., 1950, *Phys. Rev.*, **80,** 1095.

BUTLER, S. T., 1950, *Proceedings of the Harwell Nuclear Physics Conference* (Min. of Supply, Harwell), p. 54; 1951, *Proc. Roy. Soc.* A, **208,** 559.

COHEN, B. L., and FALK, C. E., 1951, *Phys. Rev.*, **84,** 173.

EL BEDEWI, F. A., MIDDLETON, R., and TAI, C. T., 1951 a, *Proc. Phys. Soc.* A, **64,** 756 ; 1951 b, *Ibid.*, **64,** 1055.

EL BEDEWI, F. A., 1952, *Proc. Phys. Soc.* A, **65,** 64.

GAMOW, G., and CRITCHFIELD, C. L., 1949, *Theory of Atomic Nucleus and Nuclear Energy Sources* (Oxford, Clarendon Press).

GOVE, H. E., 1951, *Phys. Rev.*, **81,** 364.

HOLT, J. R., and YOUNG, C. T., 1950, *Proc. Phys. Soc.* A, **63,** 833.

HUBY, R., 1950, *Nature, Lond.*, **166,** 552.

HUBY, R., and NEWNS, H. C., 1951, *Phil. Mag.*, **42,** 1442.

HUGHES, J., 1951, *Proc. Phys. Soc.* A, **64,** 797.

OPPENHEIMER, J. R., and PHILLIPS, M., 1935, *Phys. Rev.*, **48,** 500.

SERBER, R., 1947, *Phys. Rev.*, **72,** 1008.

WOLFENSTEIN, L., 1951, *Phys. Rev.*, **82,** 690.

XLVI. *On the Inhibition of Convection by a Magnetic Field*

By S. CHANDRASEKHAR, F.R.S.
Yerkes Observatory, University of Chicago*

[Received January 24, 1952]

SUMMARY

In this paper the theory of the stability of a layer of fluid heated below which has been developed by Rayleigh, Jeffreys, Pellew and Southwell is extended to the case when the fluid considered is an electrical conductor and an external magnetic field is impressed on the fluid. The problem is first investigated on the assumption that the principle of the exchange of stabilities (i.e. the equations governing marginal stability are those which are obtained from the general time dependent equations by setting $\partial/\partial t=0$) is valid. A differential equation of order six for the normal component of the velocity is derived. Suitable boundary conditions are formulated; they depend on whether the bounding surfaces are free or rigid. The case when the magnetic field and gravity act in the same direction is investigated in some detail. A variational procedure for solving the relevant equations satisfying the necessary boundary conditions and determining the critical Rayleigh numbers for the onset of convection are described. Tables of the critical Rayleigh numbers are provided for the three cases (i) both bounding surfaces free, (ii) both bounding surfaces rigid and (iii) one bounding surface free and the other rigid. Finally, the principle of the exchange of stabilities is examined and it is shown that it is valid if $4\pi\mu\sigma\kappa<1$, where μ is the magnetic permeability, σ the electrical conductivity and κ the thermometric conductivity. This condition is satisfied under most terrestial conditions; but under astrophysical conditions it is not. In the latter case, it is shown that instability can arise either as cellular convection or as over-stability (i.e. by oscillations of increasing amplitude) depending on the magnitude of the velocity of the magneto-hydrodynamic wave $(=(\mu H^2/4\pi\rho_0)^{1/2})$.

§1. INTRODUCTION

THE manner of the onset of convection in an incompressible fluid by thermal instability has been the subject of investigations by Rayleigh (1916), Jeffreys (1926, 1928), Low (1929), Pellew and Southwell (1940) and others. The principal results established by these investigations are (i) that a layer of fluid heated below first becomes unstable when the Rayleigh number

$$R=\frac{g\alpha|\beta|}{\kappa\nu}d^4, \qquad \ldots \ldots \ldots \ldots (1)$$

---

*Communicated by the Author.

(where g denotes the acceleration due to gravity, d the depth of the layer, $\beta=-|\beta|$ is the adverse temperature gradient which is maintained and α, κ and ν are the coefficients of volume expansion, thermometric conductivity and kinematic viscosity, respectively) exceeds a certain, determinate, critical value of the order of a thousand and (ii) that the motions which ensue on surpassing the critical value must have a cellular pattern in agreement with what had been demonstrated, experimentally, by Bénard (1900, 1901). The exact value of the Rayleigh number at which instability first sets in depends on boundary conditions such as whether the plane surfaces between which the layer of fluid is confined are free or rigid; and the values which have been derived theoretically have been confirmed by the experiments of Schmidt and Milverton (1935) (in this connection see also Sutton 1950).

While the criterion for instability established by the methods of Rayleigh and Jeffreys is derived explicitly for an incompressible fluid, Jeffreys (1930) has shown that the same criterion can also be applied to compressible fluids provided the density does not vary greatly within the system and provided also we interpret $|\beta|$ as the temperature gradient in excess of the adiabatic. Moreover, it follows from the discussion of Rayleigh and Jeffreys that when treating an incompressible fluid in which the temperature is variable, we can often neglect the variations of density (caused by thermal expansion) except in so far as they modify gravity: this is a method of approximation which apparently goes back to Boussinesq (1903).

Now the occurrence of magnetic fields in a number of geophysical and astrophysical problems in which convection is presumed to play a role makes it of interest to examine the effect of an external magnetic field on the onset of convection by thermal instability. It would appear that a profitable beginning in this direction may be made by re-examining the Rayleigh–Jeffreys problem for the case when the fluid considered is an electrical conductor and an external magnetic field is impressed on the fluid. We shall then have to add Maxwell's equations to the equations of hydrodynamics and heat conduction. However, it will appear that the addition of Maxwell's equations to the other equations of the problem does not introduce any essential element of difficulty; indeed, we shall see that the problem can be solved to the same degree of completeness that the classical Rayleigh–Jeffreys problem has been solved.

After this paper was completed for the press, the December 1951 issue of the *Philosophical Magazine* was received at the Yerkes Observatory. This issue contains a paper by W. B. Thompson (1951) in which the problem to which the present paper is devoted is also considered. But a comparison of the two papers will show that except for the basic idea there is very little overlapping between them: thus in the present paper the mathematical problem is treated with due attention to boundary conditions (which are effectively ignored by Thompson) and tables of the critical Rayleigh numbers for a wide range of possibilities are presented. The limits of validity of the principle of ‘exchange of stabilities’ are discussed in both papers; but it would appear that the present paper goes somewhat further than Thompson's: in particular we include the effect of viscosity which is ignored in this connection by Thompson but which turns out to be rather important for astrophysical applications.

§2. The Equations of the Problem

The equations of motion, of Maxwell, and of heat conduction appropriate to the problem on hand are:

$$\rho\frac{\partial \mathbf{u}}{\partial t}+\rho(\mathbf{u}\,.\,\mathrm{grad})\mathbf{u}-\mu\mathbf{J}\times\mathbf{H}=-\mathrm{grad}\,p+\rho\nu\nabla^2\mathbf{u}+\rho\,\mathrm{grad}\,V, \quad . \quad (2)$$

$$\left.\begin{aligned}&\mathrm{curl}\,\mathbf{H}=4\pi\mathbf{J},\ \ \mathrm{curl}\,\mathbf{E}=-\mu\frac{\partial \mathbf{H}}{\partial t},\\ &\mathrm{div}\,\mathbf{H}=0,\ \ \mathbf{J}=\sigma(\mathbf{E}+\mu\mathbf{u}\times\mathbf{H})\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$\frac{\partial T}{\partial t}+(\mathbf{u}\,.\,\mathrm{grad})T=\kappa\nabla^2 T, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where ρ denotes the density, p the pressure, T the temperature, V the gravitational potential, $\mathbf{E}$ the intensity of the electric field, $\mathbf{H}$ the intensity of the magnetic field, $\mathbf{J}$ the current density, μ the magnetic permeability and ν, κ and σ are the coefficients of kinematic viscosity, thermometric conductivity and electrical conductivity, respectively. It will be noticed that in writing Maxwell's equations we have ignored the terms in the displacement currents; as is well known this is a permissible approximation in problems of magneto-hydrodynamics. (It may also be noted here that we are adopting the electromagnetic system of units.)

Substituting curl $\mathbf{H}/4\pi$ for $\mathbf{J}$ in the equation of motion (2) and making use of the solenoidal property of $\mathbf{H}$, we can reduce it to the form

$$\rho\frac{\partial u_i}{\partial t}+\rho\frac{\partial}{\partial x_j}u_iu_j-\frac{\mu}{4\pi}\frac{\partial}{\partial x_j}H_iH_j=-\frac{\partial}{\partial x_i}\left(p+\frac{\mu|\mathbf{H}|^2}{8\pi}\right)$$
$$+\rho\nu\nabla^2u_i+\rho\frac{\partial V}{\partial x_i}, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where here and in the sequel summation over repeated indices is to be understood. Following Rayleigh we shall allow in eqns. (5) for the variation of density only in so far as it modifies the effect of the external field. Thus, in eqn. (5) we replace ρ which occurs in front of $\partial V/\partial x_i$ by

$$\rho=\rho_0(1-\alpha\Delta T) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where α denotes the coefficient of volume expansion, ρ_0 the density corresponding to a certain mean temperature T_0 and ΔT is the deviation of the local temperature from T_0:

$$\Delta T=T-T_0; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and we regard ρ occurring elsewhere in eqn. (5) as a constant equal to ρ_0. On these assumptions eqn. (5) becomes

$$\frac{\partial u_i}{\partial t}+\frac{\partial}{\partial x_j}(u_i u_j)-\frac{\mu}{4\pi\rho_0}\frac{\partial}{\partial x_j}H_i H_j=\nu\nabla^2 u_i-\alpha\Delta T\frac{\partial V}{\partial x_i}$$
$$-\frac{\partial}{\partial x_i}\left(\frac{p}{\rho_0}+\frac{\mu|\mathbf{H}|^2}{8\pi\rho_0}-V\right). \quad . \quad . \quad (8)$$

With the variation of density due to thermal expansion allowed for in this manner, we treat u_i from now on as a solenoidal vector:

$$\frac{\partial u_i}{\partial x_i}=0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

When the external field is that due to the action of gravity, we can write

$$\frac{\partial V}{\partial x_i}=-g\lambda_i \text{ and } V=-g\lambda_j x_j, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where $\boldsymbol{\lambda}$ is a unit vector in the direction of the vertical. With the foregoing substitution, eqn. (8) becomes

$$\frac{\partial u_i}{\partial t}+\frac{\partial}{\partial x_j}u_i u_j-\frac{\mu}{4\pi\rho_0}\frac{\partial}{\partial x_j}H_i H_j=\nu\nabla^2 u_i+\gamma\Delta T\lambda_i$$
$$-\frac{\partial}{\partial x_i}\left(\frac{p}{\rho_0}+\frac{\mu|\mathbf{H}|^2}{8\pi\rho_0}+g\lambda_j x_j\right), \quad . \quad . \quad . \quad (11)$$

where we have written

$$\gamma=g\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Turning to the equation of heat conduction (4), we shall suppose that a constant mean temperature gradient $\beta=-|\beta|$ is maintained in the direction of $\boldsymbol{\lambda}$ by an external agency and that we can write

$$T=T_0+\beta\lambda_j x_j+\theta,\ \Delta T=\beta\lambda_j x_j+\theta \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where T_0 is a constant (cf. eqn. (7)) and

$$\beta=\frac{\partial T}{\partial x_j}\lambda_j, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

and θ is the deviation of the temperature from its local mean value $T_0+\beta\lambda_j x_j$. Inserting (13) in eqns. (4) and (11) and remembering that H_i is solenoidal, we obtain

$$\frac{\partial u_i}{\partial t}+\frac{\partial}{\partial x_j}u_i u_j-\frac{\mu}{4\pi\rho_0}H_j\frac{\partial H_i}{\partial x_j}=\nu\nabla^2 u_i+\gamma\theta\lambda_i-\frac{\partial\varpi}{\partial x_i} \quad . \quad . \quad . \quad (15)$$

and

$$\frac{\partial\theta}{\partial t}+\beta\lambda_j u_i+u_j\frac{\partial\theta}{\partial x_i}=\kappa\nabla^2\theta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where, for brevity, we have written

$$\varpi=\frac{p}{\rho_0}+\frac{\mu|\mathbf{H}|^2}{8\pi\rho_0}+g\lambda_j x_j-\tfrac{1}{2}\beta\gamma\lambda_i\lambda_j x_i x_j. \quad . \quad . \quad . \quad . \quad (17)$$

Considering next Maxwell's eqns. (3), we obtain for $\mathbf{H}$ the equation

$$\mu\frac{\partial \mathbf{H}}{\partial t}=-\text{curl}\left(\frac{\mathbf{J}}{\sigma}-\mu\mathbf{u}\times\mathbf{H}\right)=\mu\,\text{curl}\,(\mathbf{u}\times\mathbf{H})-\frac{1}{\sigma}\,\text{curl}\,\mathbf{J}$$

$$=\mu\,\text{curl}\,(\mathbf{u}\times\mathbf{H})+\frac{1}{4\pi\sigma}\,\nabla^2\mathbf{H}. \qquad (18)$$

Hence,

$$\frac{\partial \mathbf{H}}{\partial t}-\text{curl}\,(\mathbf{u}\times\mathbf{H})=\eta\nabla^2\mathbf{H}, \qquad (19)$$

where

$$\eta=\frac{1}{4\pi\mu\sigma}. \qquad (20)$$

Using the solenoidal property of u_i and H_i, we can rewrite eqn. (19) alternatively in the form

$$\frac{\partial H_i}{\partial t}+u_j\frac{\partial H_i}{\partial x_j}-H_j\frac{\partial u_i}{\partial x_j}=\eta\nabla^2 H_i. \qquad (21)$$

If a constant magnetic field $H=|\mathbf{H}|$ is maintained, externally, in the direction of a unit vector $\boldsymbol{\tau}$, we may write

$$H_i=H\tau_i+h_i, \qquad (22)$$

where h_i is the perturbation in the magnetic field caused by the motions and the induced currents in the fluid.

We shall now suppose that u_i, θ and h_i are small quantities of the first order and that we can ignore products and squares of them. On this approximation, eqns. (15), (16) and (21) reduce to

$$\frac{\partial u_i}{\partial t}=\frac{\mu H}{4\pi\rho_0}\tau_j\frac{\partial h_i}{\partial x_j}+\gamma\theta\lambda_i+\nu\nabla^2 u_i-\frac{\partial \varpi}{\partial x_i}, \qquad (23)$$

$$\frac{\partial \theta}{\partial t}=-\beta\lambda_j u_j+\kappa\nabla^2\theta, \qquad (24)$$

and

$$\frac{\partial h_i}{\partial t}=H\tau_j\frac{\partial u_i}{\partial x_j}+\eta\nabla^2 h_i. \qquad (25)$$

Taking the divergence of eqn. (23) and remembering that u_i and h_i are both solenoidal, we find that

$$\nabla^2\varpi=\gamma\lambda_j\frac{\partial \theta}{\partial x_j}. \qquad (26)$$

In the further discussion of the foregoing equations, we shall find it convenient to work with the velocity, w, and the intensity of the magnetic field, h, in the direction of $\boldsymbol{\lambda}$. Thus, letting

$$w=u_i\lambda_i \text{ and } h=h_i\lambda_i, \qquad (27)$$

we readily derive from eqns. (23) to (25) the following set of equations:

$$\frac{\partial w}{\partial t}=\frac{\mu H}{4\pi\rho_0}\tau_j\frac{\partial h}{\partial x_j}+\gamma\theta+\nu\nabla^2 w-\lambda_i\frac{\partial \varpi}{\partial x_i}, \qquad (28)$$

$$\frac{\partial \theta}{\partial t}=-\beta w+\kappa\nabla^2\theta, \qquad (29)$$

and

$$\frac{\partial h}{\partial t}=H\tau_j\frac{\partial w}{\partial x_j}+\eta\nabla^2 h. \qquad (30)$$

These equations together with eqn. (26) represent the basic equations of this theory.

As stated in § 1, we shall consider a horizontal layer of fluid of depth, d, confined between two parallel planes at $z=0$ and $z=d$. The conditions to be satisfied on these bounding surfaces will depend on whether they are free or rigid. But in all cases, we must require that

$$\theta=w=0 \text{ for } z=0 \text{ and } z=d, \qquad (31)$$

since (by assumption) the temperatures on these surfaces are maintained constant and there can also be no motion normal to them. Additional boundary conditions follow from the equation of continuity (9) and depend on whether the bounding surfaces are free or rigid; they are (cf. Pellew and Southwell 1940):

$$\frac{\partial^2 w}{\partial z^2}=0 \quad \text{on a free surface} \qquad (32)$$

and

$$\frac{\partial w}{\partial z}=0 \quad \text{on a rigid surface.} \qquad (33)$$

§ 3. THE EQUATIONS GOVERNING MARGINAL STABILITY

If the principle of the exchange of stabilities as described by Jeffreys (1926) is valid also for the problem under consideration, then we can write down at once the equations governing the situation in marginal stability (i.e. one on the verge of stability) by setting $\partial/\partial t=0$ in eqns. (28) to (30). In § 8 we shall examine the validity of this principle for the problem on hand; and it will appear from the discussion there that the principle will be valid under most terrestial conditions in which we shall want to apply the theory. We shall therefore accept the principle for the present, postponing to § 8 an examination of the limits of its validity.

Setting $\partial/\partial t=0$ in eqns. (28) to (30) in accordance with the principle of the exchange of stabilities, we obtain the following equations governing marginal stability:

$$\nu\nabla^2 w=\lambda_i\frac{\partial \varpi}{\partial x_i}-\gamma\theta-\frac{\mu H}{4\pi\rho_0}\tau_j\frac{\partial h}{\partial x_j}, \qquad (34)$$

$$\kappa\nabla^2\theta=\beta w, \qquad (35)$$

$$\eta\nabla^2 h=-H\tau_j\frac{\partial w}{\partial x_j}, \qquad (36)$$

and

$$\nabla^2\varpi=\gamma\lambda_j\frac{\partial\theta}{\partial x_j}. \qquad (37)$$

Operating eqn. (34) by ∇^2 and making use of eqns. (20), (36) and (37), we obtain

$$\nabla^4 w=\frac{\gamma}{\nu}\left(\lambda_i\lambda_j\frac{\partial^2\theta}{\partial x_i\partial x_j}-\nabla^2\theta\right)+\frac{\mu^2H^2\sigma}{\rho_0\nu}\tau_i\tau_j\frac{\partial^2 w}{\partial x_i\partial x_j}. \qquad (38)$$

Operating this equation once again by ∇^2 and making use of eqn. (35), we have

$$\nabla^6 w = \frac{|\beta|\gamma}{\kappa\nu}\left(\nabla^2 w - \lambda_i\lambda_j\frac{\partial^2 w}{\partial x_i \partial x_j}\right) + \frac{\mu^2 H^2 \sigma}{\rho_0 \nu}\tau_i\tau_j\frac{\partial^2}{\partial x_i \partial x_j}\nabla^2 w. \quad . \quad (39)$$

It is also evident that θ satisfies an equation of exactly the same form as w.

§4. THE REDUCTION TO A BOUNDARY VALUE PROBLEM WHEN BOTH H AND g ACT IN THE SAME DIRECTION

In some ways the case of greatest interest is when the external magnetic field and gravity both act in the same direction. If this common direction be chosen as the z-direction, then

$$\boldsymbol{\lambda} = \boldsymbol{\tau} = (0, 0, 1), \quad . \; . \; . \; . \; . \; . \; . \; . \quad (40)$$

and eqns. (38) and (39) of the preceding section reduce to

$$\nabla^4 w = -\frac{\gamma}{\nu}\left(\frac{\partial^2\theta}{\partial x^2} + \frac{\partial^2\theta}{\partial y^2}\right) + \frac{\mu^2 H^2 \sigma}{\rho_0 \nu}\frac{\partial^2 w}{\partial z^2}, \quad . \; . \; . \; . \quad (41)$$

and

$$\nabla^6 w = \frac{|\beta|\gamma}{\kappa\nu}\left(\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2}\right) + \frac{\mu^2 H^2 \sigma}{\rho_0 \nu}\frac{\partial^2}{\partial z^2}\nabla^2 w. \quad . \; . \; . \quad (42)$$

In analogy with the classical Rayleigh–Jeffreys problem, we may expect that in the present case also, convection when it first sets in will have a cellular pattern with the cell walls being surfaces of symmetry. We may then separate the variables x and y from z by assuming that w and θ have the forms

$$w = f(x, y)W(z) \text{ and } \theta = f(x, y)\Theta(z), \quad . \; . \; . \; . \quad (43)$$

where

$$d^2\left(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right) = -a^2 f, \quad . \; . \; . \; . \; . \; . \; . \quad (44)$$

and a^2 is a certain ‘ characteristic number ’ determined by the conditions on the cell walls. And if as we have assumed, the cell walls are surfaces of symmetry, we must require that the normal gradient $\partial f/\partial n$ vanishes on the cell walls. As Pellew and Southwell (1940) have shown this boundary condition on f ensures the reality of a.

The solution for f when the boundary of the cell is a rectangle of sides L_1 and L_2 is, clearly

$$f = A_{mn}\cos\frac{m\pi x}{L_1}\cos\frac{n\pi y}{L_2}, \quad . \; . \; . \; . \; . \; . \; . \quad (45)$$

where m and n are arbitrary even integers and A_{mn} is a constant; the corresponding relation between a, L_1, L_2, and the depth of the cell, d, is

$$a^2 = d^2\pi^2\left(\frac{m^2}{L_1^2} + \frac{n^2}{L_2^2}\right). \quad . \; . \; . \; . \quad (46)$$

The solution of eqn. (44) when the boundary of the cell is a regular hexagon with a side L has been discovered by Christopherson (1940). His solution is

$$f=f_0\left\{\cos\frac{2n\pi}{3L}(\sqrt{3}x+y)+\cos\frac{2n\pi}{3L}(\sqrt{3}x-y)+\cos\frac{4n\pi y}{3L}\right\}, \quad . \quad (47)$$

where f_0 is a constant and n is an integer; and the equation relating a, L and d is (cf. Pellew and Southwell 1940)

$$a=\left(\frac{4n\pi}{3L}\right)d. \quad . \; . \; . \; . \; . \; . \; . \quad (48)$$

Returning to eqns. (41) and (42), substituting for w and θ in accordance with eqns. (43) and (44), and letting

$$\zeta=zd \text{ and } D=\frac{d}{d\zeta}, \quad . \; . \; . \; . \; . \; . \; . \quad (49)$$

we have

$$[(D^2-a^2)^2-QD^2]W=\frac{\gamma}{\nu}a^2d^2\Theta, \quad . \; . \; . \; . \; . \; . \quad (50)$$

and

$$(D^2-a^2)[(D^2-a^2)^2-QD^2]W=-a^2RW, \quad . \; . \; . \quad (51)$$

where

$$Q=\frac{\mu^2H^2\sigma}{\rho_0\nu}d^2 \text{ and } R=\frac{|\beta|\gamma}{\kappa\nu}d^4. \quad . \; . \; . \; . \; . \; . \quad (52)$$

The boundary conditions with respect to which eqns. (50) and (51) must be solved are (cf. eqns. (31) to (33)):

$$\Theta=W=0 \text{ for } \zeta=0 \text{ and } 1, \quad . \; . \; . \; . \; . \; . \; . \quad (53)$$

and

$$\left.\begin{aligned} DW&=0 \quad \text{on a rigid surface,}\\ D^2W&=0 \quad \text{on a free surface.}\end{aligned}\right\} \quad . \; . \; . \; . \; . \quad (54)$$

In view of eqn. (50), the boundary conditions (53) are equivalent to

$$W=[(D^2-a^2)^2-QD^2]W=0 \text{ for } \zeta=0 \text{ and } 1. \quad . \; . \; . \quad (55)$$

On a free surface the boundary conditions (54) and (55) can be combined to give

$$W=D^2W=D^4W=0 \quad \text{(on a free surface).} \quad . \; . \; . \quad (56)$$

Equations (54) and (55) provide six boundary conditions and the requirement that a solution of eqn. (51) satisfies these conditions will determine for any given a^2 a sequence of possible values for R. For, eqn. (51) being a linear equation with constant coefficients of order six, its general solution must be of the form

$$W=\sum_{i=1}^{6}A_i e^{q_i\zeta}, \quad . \; . \; . \; . \; . \; . \; . \; . \quad (57)$$

where the q_i's, occuring in pairs, are the roots of the characteristic equation

$$(q^2-a^2)[(q^2-a^2)^2-Qq^2]=-a^2R, \quad . \; . \; . \; . \quad (58)$$

and the A_i's are constants of integration. The six boundary conditions (three at $\zeta=0$ and three at $\zeta=1$) will lead to a set of six linear homogeneous equations for the A_i's and the condition that the determinant of this system vanishes (so that we may have a non-trivial solution) will determine a sequence of possible values for R. Among these possible values (for a given a^2) there will be a lowest value; the minimum of these lowest values (as a function of a^2) will finally specify the critical Rayleigh number, R_c, at which convection, through instability, will first set in.

While there is, of course, no difficulty of principle in carrying out the procedure outlined in the preceding paragraph, it is evident that the actual carrying out of the procedure for various values of Q and boundary conditions will involve a prohibitive amount of work. Thus, even for the classical Rayleigh–Jeffreys problem the precise determination of R_c for the two standard cases (i) both bounding surfaces rigid and (ii) one bounding surface rigid and the other free, was completed only in 1940 by Pellew and Southwell. However, it will appear that the extension of the variational principle established by Pellew and Southwell for the Rayleigh–Jeffreys problem to the present one leads to a convenient practical procedure for determining R_c without an undue amount of labour. The variational principle is developed in the following section.

§5. A VARIATIONAL PRINCIPLE FOR DETERMINING R_c WHEN BOTH $\mathbf{H}$ AND $\mathbf{g}$ ACT IN THE SAME DIRECTION

First, we shall obtain an integral expression for R. Letting

$$F=[(D^2-a^2)^2-QD^2]W, \qquad \ldots\ldots \quad (59)$$

we shall rewrite the differential equation governing W in the form (cf. eqn. (51))

$$(D^2-a^2)F=-a^2RW. \qquad \ldots\ldots\ldots \quad (60)$$

The boundary conditions (55) require that

$$F=W=0 \text{ for } \zeta=0 \text{ and } 1. \qquad \ldots\ldots \quad (61)$$

Now multiply eqn. (60) by F and integrate over the range of ζ (i.e. from 0 to 1). The left-hand side of the equation gives

$$\int_0^1 F(D^2-a^2)F\,d\zeta=\int_0^1 FD^2F\,d\zeta-a^2\int_0^1 F^2\,d\zeta$$

$$=[FDF]_0^1-\int_0^1 \{(DF)^2+a^2F^2\}d\zeta. \qquad \ldots\ldots \quad (62)$$

By virtue of the boundary conditions on F the integrated part vanishes and we are left with

$$\int_0^1 F(D^2-a^2)F\,d\zeta=-\int_0^1 \{(DF)^2+a^2F^2\}d\zeta. \qquad \ldots \quad (63)$$

Turning next to the right-hand side of eqn. (60), we have to consider

$$\int_0^1 WF\,d\zeta=\int_0^1 W(D^2-a^2)^2W\,d\zeta-Q\int_0^1 WD^2W\,d\zeta. \quad . \quad . \quad . \quad (64)$$

Considering the first of the two integrals on the right-hand side of (64), we obtain after two successive integrations by parts

$$\begin{aligned}&\int_0^1 W(D^2-a^2)[(D^2-a^2)W]d\zeta\\&\quad=\int_0^1 WD^2[(D^2-a^2)W]d\zeta-a^2\int_0^1 W(D^2-a^2)W\,d\zeta\\&\quad=\{WD[(D^2-a^2)W]-DW[(D^2-a^2)W]\}_0^1\\&\quad+\int_0^1 D^2W[(D^2-a^2)W]d\zeta-a^2\int_0^1 W(D^2-a^2)W\,d\zeta. \quad . \quad . \quad (65)\end{aligned}$$

The integrated terms vanish on account of the boundary condition $W=0$ and either of (54). Thus,

$$\int_0^1 W(D^2-a^2)^2W\,d\zeta=\int_0^1 [(D^2-a^2)W]^2d\zeta. \quad . \quad . \quad . \quad (66)$$

On the other hand

$$\int_0^1 WD^2W\,d\zeta=[WDW]_0^1-\int_0^1 (DW)^2d\zeta=-\int_0^1 (DW)^2d\zeta. \quad . \quad (67)$$

Combining (64), (66) and (67), we have

$$\int_0^1 WF\,d\zeta=\int_0^1 \{[(D^2-a^2)W]^2+Q(DW)^2\}d\zeta. \quad . \quad . \quad . \quad (68)$$

The result of multiplying eqn. (60) by F and integrating over ζ is, therefore,

$$R=\frac{\int_0^1 \{(DF)^2+a^2F^2\}d\zeta}{a^2\int_0^1 \{[(D^2-a^2)W]^2+Q(DW)^2\}d\zeta}=\frac{I_1}{a^2I_2}\text{ (say)}. \quad . \quad . \quad (69)$$

This formula expresses R as the ratio of two positive definite integrals.

Consider now the effect on R of an arbitrary variation δW in W compatible only with the boundary conditions on W. We have, to the first order

$$\delta R=\frac{1}{a^2I_2}\left(\delta I_1-\frac{I_1}{I_2}\delta I_2\right)=\frac{1}{a^2I_2}(\delta I_1-a^2R\delta I_2), \quad . \quad . \quad (70)$$

where δI_1 and δI_2 denote the corresponding variations in I_1 and I_2:

$$\left.\begin{aligned}&\delta I_1=2\int_0^1 \{(DF)(D\delta F)+a^2F\delta F\}d\zeta\\&\text{and}\\&\delta I_2=2\int_0^1 \{[(D^2-a^2)W][(D^2-a^2)\delta W]+Q(DW)(D\delta W)\}d\zeta.\end{aligned}\right\} \quad . \quad (71)$$

The expressions for δI_1 and δI_2 can be reduced by integration by parts and making use of the boundary conditions

$$\left.\begin{array}{ll} \delta W=\delta F=0 \text{ for } \zeta=0 \text{ and } 1 & \\ D^2\delta W=0 & \text{on a free surface, and} \\ D\delta W=0 & \text{on a rigid surface.} \end{array}\right\} \quad . \quad . \quad (72)$$

Thus

$$\delta I_1=2[\delta F\, DF]_0^1-2\int_0^1 \delta F(D^2-a^2)F\,d\zeta=-2\int_0^1 \delta F(D^2-a^2)F\,d\zeta \quad . \quad (73)$$

and

$$\begin{aligned} \delta I_2= & 2\{DW[(D^2-a^2)\delta W]-WD[(D^2-a^2)\delta W]+QWD\delta W\}_0^1 \\ & +2\int_0^1 W\{[(D^2-a^2)^2-QD^2]\delta W\}d\zeta \\ = & 2\int_0^1 W\delta F\,d\zeta. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (74) \end{aligned}$$

Now combining eqns. (70), (73) and (74), we obtain

$$\delta R=-\frac{2}{a^2 I_2}\int_0^1 \delta F\{(D^2-a^2)F+a^2RW\}d\zeta. \quad . \quad . \quad . \quad (75)$$

Hence, to the first order $\delta R=0$ for all small arbitrary variations in W satisfying the boundary conditions provided

$$(D^2-a^2)F+a^2RW=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

i.e. if the differential equation governing W is satisfied. It is evident that the converse of this proposition is also true. Further, it follows from (75) that the true solution of the problem leads to a minimum value for R when evaluated according to formula (69). This last fact enables us to formulate the following variational procedure of solving eqn. (51) (for any assigned a^2) and satisfying the boundary conditions of the problem.

Assume for F an expression involving one or more parameters, A_k, and which vanishes at $\zeta=0$ and 1. With the chosen form of F determine W as a solution of the equation

$$[(D^2-a^2)^2-QD^2]W=F, \quad . \quad . \quad . \quad . \quad . \quad . \quad (77)$$

and satisfying the boundary conditions on W at $\zeta=0$ and 1. Then evaluate R according to formula (69) and minimize it with respect to the parameters A_k. In this way, we shall obtain the 'best' value of R for the chosen form of F. We shall see in the following section that even with the simplest trial functions we reach quite high accuracy in the deduced values of R.

§6. THE DETERMINATION OF R_c WHEN $\mathbf{H}$ AND $\mathbf{g}$ ACT IN THE SAME DIRECTION AND FOR THE CASES (i) WHEN BOTH BOUNDING SURFACES ARE FREE, (ii) WHEN BOTH BOUNDING SURFACES ARE RIGID AND (iii) WHEN ONE BOUNDING SURFACE IS FREE AND THE OTHER IS RIGID

(i) *Both bounding surfaces free.*—In this case the solution of eqn. (51) satisfying the boundary conditions (cf. eqn. (56))

$$W=D^2W=D^4W=0 \text{ for } \zeta=0 \text{ and } 1, \qquad (78)$$

can be explicitly written down. Thus

$$W=\text{constant} \sin n\pi\zeta, \qquad (79)$$

where n is an integer clearly satisfies the boundary conditions (78) and represents a solution of eqn. (51) provided

$$R=\frac{1}{a^2}(n^2\pi^2+a^2)[(n^2\pi^2+a^2)^2+Qn^2\pi^2]. \qquad (80)$$

That the solution obtained in this manner by inspection is in fact the unique solution of the problem can be confirmed by verifying that formula (69) also leads to the result (80), if the choice

$$F=A \sin n\pi\zeta \quad (A=\text{constant}), \qquad (81)$$

be made for the trial function and the corresponding solution of eqn. (77) which satisfies the boundary conditions (75), namely

$$W=\frac{A \sin n\pi\zeta}{(n^2\pi^2+a^2)^2+Qn^2\pi^2}, \qquad (82)$$

be used in the evaluation of R.

Letting $a^2=n^2\pi^2x$, we can rewrite eqn. (80) in the form

$$R=n^4\pi^4\frac{1+x}{x}\left[(1+x)^2+\frac{Q}{n^2\pi^2}\right]. \qquad (83)$$

From this equation it follows that for a given x (i.e. a^2) instability first sets in for the lowest mode $n=1$. The corresponding expression for R is

$$R=\pi^4\frac{1+x}{x}[(1+x)^2+Q_1] \quad \left(Q_1=\frac{Q}{\pi^2}\right). \qquad (84)$$

As a function of x, R given by the foregoing formula attains its minimum when

$$2x^3+3x^2=Q_1+1. \qquad (85)$$

With x determined as a solution of this cubic equation, (84) will give the critical Rayleigh number, R_c, at which convection through instability will first set in. The values of R_c determined in this fashion for various values of Q are given in table 1. The inhibiting effect of a magnetic field on the onset of convection is apparent from this table.

For Q_1 sufficiently large the root of eqn. (85) tends to

$$x_{\min}\rightarrow(Q_1/2)^{1/3}; \qquad (86)$$

and the corresponding limiting expression for R_c is

$$R_c=\pi^4[Q_1+(\tfrac{1}{2}Q_1)^{2/3}]. \qquad (87)$$

It is found that this last formula for R_c gives sufficient accuracy for $Q>1\,000$.

Table 1. Critical Rayleigh Numbers for the Case when both Bounding Surfaces are Free

Q	a	R_c	R_c/R_o	Q	a	R_c	R_c/R_o
0	2·233	657·511	1·0000	6500	7·952	80343·6	122·19
5	2·432	796·573	1·2115	7000	8·059	86034·0	130·85
10	2·590	923·070	1·4039	7500	8·159	91705·7	139·47
20	2·826	1154·19	1·7554	8000	8·253	97360·1	148·07
50	3·270	1762·04	2·6799	8500	8·343	102998·0	156·65
100	3·702	2653·71	4·0360	9000	8·429	108623·0	165·20
150	3·990	3475·67	5·2861	9500	8·510	114234·0	173·74
200	4·210	4258·49	6·4767	10000	8·588	119832·0	182·25
300	4·543	5752·65	8·7491	10500	8·663	125419·0	190·75
400	4·794	7185·94	10·929	11000	8·735	130995·0	199·23
500	4·998	8578·88	13·048	11500	8·804	136560·0	207·69
600	5·171	9942·40	15·121	12000	8·870	142116·0	216·14
700	5·321	11283·2	17·160	13000	8·997	153202·0	233·00
800	5·455	12605·6	19·172	14000	9·116	164255·0	249·81
1000	5·684	15207·0	23·128	15000	9·227	175279·0	266·58
1500	6·123	21535·2	32·753	16000	9·333	186276·0	283·30
2000	6·453	27699·9	42·128	17000	9·433	197249·0	299·99
2500	6·720	33756·5	51·340	18000	9·528	208199·0	316·65
3000	6·945	39734·2	60·431	19000	9·619	219129·0	333·27
3500	7·140	45650·6	69·429	20000	9·706	230038·0	349·86
4000	7·313	51517·8	78·353	25000	10·09	284341·0	432·45
4500	7·442	57344·6	87·215	30000	10·42	338308·0	514·53
5000	7·585	63135·9	96·022	35000	10·70	392013·0	596·21
5500	7·717	68897·3	104·78	40000	10·95	445507·0	677·56
6000	7·839	74632·1	113·51				

(ii) *Both bounding surfaces rigid.*—We shall obtain the solution for this case by an application of the variational procedure described in § 5. But first it is convenient to translate the ζ-axis so that the bounding surfaces are at $\zeta=\pm\frac{1}{2}$. With the bounding surfaces at $\zeta=\pm\frac{1}{2}$ it follows from the identity of the boundary conditions to be satisfied at $\zeta=\pm\frac{1}{2}$ that the solutions for F and W should be either even or odd functions of ζ; and it can be readily verified that the even solutions give lower values for R. We must therefore choose the even solutions.

As a simple trial function for F which satisfies the boundary condition, $F=0$ at $\zeta=\pm\frac{1}{2}$, we shall assume

$$F=A\cos\pi\zeta+1+\cos 2\pi\zeta \quad (-\tfrac{1}{2}\leqslant\zeta\leqslant+\tfrac{1}{2}), \qquad (88)$$

where A is a variational parameter. With F given by (88), the equation governing W is

$$[(D^2-a^2)^2-QD^2]W=1+\cos 2\pi\zeta+A\cos\pi\zeta\,; \qquad (89)$$

and the solution of this equation appropriate to the problem on hand is

$$W=\gamma_0+\gamma_2\cos 2\pi\zeta+A\gamma_1\cos\pi\zeta+B\cosh q_1\zeta+C\cosh q_2\zeta, \qquad (90)$$

where B and C are constants of integration,

$$q_1=\tfrac{1}{2}[\sqrt{(Q+4a^2)}+\sqrt{Q}],\ q_2=\tfrac{1}{2}[\sqrt{(Q+4a^2)}-\sqrt{Q}], \qquad (91)$$

and

$$\gamma_n=\frac{1}{(n^2\pi^2+a^2)^2+n^2\pi^2 Q}. \qquad (92)$$

The boundary conditions to be satisfied by W are (cf. eqns. (53) and (54))

$$W=DW=0 \text{ for } \zeta=\pm\tfrac{1}{2}. \qquad (93)$$

These conditions require that

$$\left.\begin{aligned} &B\cosh\tfrac{1}{2}q_1+C\cosh\tfrac{1}{2}q_2=-(\gamma_0-\gamma_2)\\ \text{and}\quad &Bq_1\sinh\tfrac{1}{2}q_1+Cq_2\sinh\tfrac{1}{2}q_2=A\pi\gamma_1.\end{aligned}\right\} \qquad (94)$$

The solutions of these equations can be expressed in the forms:

$$\left.\begin{aligned} B&=+\delta[A\pi\gamma_1+q_2(\gamma_0-\gamma_2)\tanh\tfrac{1}{2}q_2]\operatorname{sech}\tfrac{1}{2}q_1,\\ C&=-\delta[A\pi\gamma_1+q_1(\gamma_0-\gamma_2)\tanh\tfrac{1}{2}q_1]\operatorname{sech}\tfrac{1}{2}q_2,\end{aligned}\right\} \qquad (95)$$

where

$$\delta=(q_1\tanh\tfrac{1}{2}q_1-q_2\tanh\tfrac{1}{2}q_2)^{-1}. \qquad (96)$$

With F and W given by eqns. (88), (90) and (95) we find that

$$\int_{-1/2}^{+1/2}[(DF)^2+a^2F^2]d\zeta=\tfrac{1}{2}(\pi^2+a^2)\left[A^2+\frac{32}{3\pi}A+\frac{4\pi^2+3a^2}{\pi^2+a^2}\right]. \qquad (97)$$

and

$$\begin{aligned}\int_{-1/2}^{+1/2}WF\,d\zeta=\tfrac{1}{2}\gamma_1&\left[1-\frac{4\pi^2\delta(q_1{}^2-q_2{}^2)}{(\pi^2+q_1{}^2)(\pi^2+q_2{}^2)}\right]A^2\\ &+\left[\frac{2}{3\pi}(3\gamma_0+4\gamma_1+\gamma_2)-2\pi\delta(\gamma_0-\gamma_2)\left\{\frac{q_1\tanh\frac{1}{2}q_1}{\pi^2+q_2{}^2}-\frac{q_2\tanh\frac{1}{2}q_2}{\pi^2+q_1{}^2}\right\}\right.\\ &\left.-8\pi^3\gamma_1\delta\left\{\frac{\tanh\frac{1}{2}q_2}{q_2(4\pi^2+q_2{}^2)}-\frac{\tanh\frac{1}{2}q_1}{q_1(4\pi^2+q_1{}^2)}\right\}\right]A+\left[(\gamma_0+\tfrac{1}{2}\gamma_2)\right.\\ &\left.-8\pi^2(\gamma_0-\gamma_2)\delta\tanh\tfrac{1}{2}q_1\tanh\tfrac{1}{2}q_2\frac{(q_1{}^2-q_2{}^2)(4\pi^2+q_1{}^2+q_2{}^2)}{a^2(4\pi^2+q_1{}^2)(4\pi^2+q_2{}^2)}\right]. \qquad (98)\end{aligned}$$

The resulting expression for R is, therefore, of the form

$$R=\frac{\int_{-1/2}^{+1/2}[(DF)^2+a^2F^2]d\zeta}{a^2\int_{-1/2}^{+1/2}WF\,d\zeta}=\frac{f+gA+hA^2}{l+mA+nA^2}. \qquad (99)$$

As a function of A, R attains its minimum when A is a root of the equation

$$(gl-mf)+2(hl-fn)A+(hm-gn)A^2=0. \quad . \quad . \quad . \quad (100)$$

With A determined by this equation, eqn. (99) determines the 'best' value of R for any given a. By repeating the calculations for various assigned values of a we can determine the value of R_c. However, in practise it was found more convenient to carry out these calculations in two stages in the manner described below.

First, we may observe that if we had assumed for F the trial function

$$F=A\cos\pi\zeta, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (101)$$

we should have found for R the formula

$$R=\frac{(\pi^2+a^2)[(\pi^2+a^2)^2+\pi^2Q]}{a^2[1-4\pi^2\delta(q_1{}^2-q_2{}^2)/(\pi^2+q_1{}^2)(\pi^2+q_2{}^2)]}, \quad . \quad . \quad . \quad (102)$$

which follows from eqns. (97) to (99) by letting $A\rightarrow\infty$. It appears that this formula which has no variational parameter in it already determines R to an accuracy better than 1%. For in the field-free case when $Q=0$ and $q_1=q_2=a$ the limiting form of (102), namely

$$R=\frac{(\pi^2+a^2)^3}{a^2[1-1{\cdot}6\pi^2a\cosh^2\frac{1}{2}a/(a+\sinh a)(\pi^2+a^2)^2]}, \quad . \quad . \quad (103)$$

gives $R_c=1\,715{\cdot}1$ for $a=3{\cdot}13$; and this value should be compared with $R_c=1\,707{\cdot}8$ determined by Pellew and Southwell (1940) by an exact solution of the problem. The procedure then was the following:

For an assigned Q the value of a which gave the minimum value for R evaluated according to eqn. (102) was first determined. With this value of a a 'second approximation' was then carried out with the aid of eqns. (97) to (99) by making the variation with respect to A. The results of such calculations are summarized in table 2. The extent of the agreement between the results of the first and the second approximations may be considered as a measure of the accuracy of the determination

Table 2. Critical Rayleigh Numbers and Related Constants for the Case when both Bounding Surfaces are Rigid

Q	a	A	First Approximation	Second Approximation	R_c/R_0
0	3·13		1715·1	1707·8	1·0000
10	3·25	10·506	1953·7	1945·9	1·1394
50	3·68	10·963	2811·4	2802·1	1·6408
100	4·00	11·687	3767·6	3757·4	2·2001
200	4·45	12·910	5499·9	5488·6	3·2138
500	5·16	15·917	10122·0	10110·0	5·9199
1000	5·80	19·261	17116·0	17103·0	10·015
2000	6·55	23·830	30139·0	30125·0	17·640
4000	7·40	29·896	54711·0	54697·0	32·028
6000	7·94	34·500	78405·0	78391·0	45·902
8000	8·34	38·052	101622·0	101606·0	59·495
10000	8·66	40·462	124523·0	124509·0	72·906

of R_c by the variational method. In this connection it should be noted that an uncertainty in a of 0·1 in the neighbourhood where R_c attains its minimum does not affect the deduced values of R by more than one part in 10^5; and often, it is much less.

Finally, we may remark that the asymptotic behaviour of R for large Q is the same as in case (i).

(iii) *One bounding surface free and the other rigid.*—In this case the conditions to be satisfied on the two bounding surfaces are different. But as has already been remarked by Low (1929) and Pellew and Southwell (1940) the solution for this case can be reduced to that of case (ii) by considering the odd (instead of the even) solutions. For, it is evident that solutions for F and W which are odd and satisfy the boundary conditions of case (ii) at $\zeta=\pm\frac{1}{2}$, vanish at $\zeta=0$ and satisfy here the boundary conditions appropriate to a free surface. Consequently, an odd solution for case (ii) applicable to a cell depth d also provides a solution for case (iii) and applicable to a cell depth $d'=\frac{1}{2}d$ and a Rayleigh number which is 16 times smaller.

In accordance with the foregoing remarks we shall assume for F the trial function

$$F=A\sin 2\pi\zeta+\sin\pi\zeta+\sin 3\pi\zeta. \qquad . \quad . \quad . \qquad . \quad (104)$$

The corresponding solution for W is (cf. eqn. (90))

$$W=\gamma_1\sin\pi\zeta+\gamma_3\sin 3\pi\zeta+A\gamma_2\sin 2\pi\zeta+B\sinh q_1\zeta+C\sinh q_2\zeta,$$
$$. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (105)$$

where γ_n, q_1 and q_2 have the same meanings as in eqns. (91) and (92). And the integration constants B and C determined by the boundary conditions (93) are:

$$\left.\begin{aligned} B&=+\delta'[2\pi\gamma_2A+q_2(\gamma_1-\gamma_3)\coth\tfrac{1}{2}q_2]\operatorname{cosech}\tfrac{1}{2}q_1,\\ C&=-\delta'[2\pi\gamma_2A+q_1(\gamma_1-\gamma_3)\coth\tfrac{1}{2}q_1]\operatorname{cosech}\tfrac{1}{2}q_2,\end{aligned}\right\} \quad . \quad . \quad (106)$$

where

$$\delta'=[q_1\coth\tfrac{1}{2}q_1-q_2\coth\tfrac{1}{2}q_2]^{-1}. \qquad . \quad . \quad . \quad . \quad (107)$$

With F and W given by the foregoing equations we find that

$$\int_{-1/2}^{+1/2}[(DF)^2+a^2F^2]d\zeta=\tfrac{1}{2}(4\pi^2+a^2)\left[A^2+\frac{128}{15\pi}A+\frac{2(5\pi^2+a^2)}{4\pi^2+a^2}\right] \qquad (108)$$

and

$$\begin{aligned}\int_{-1/2}^{+1/2}WF\,d\zeta=\tfrac{1}{2}\gamma_2\left[1-\frac{16\pi^2\delta'(q_1^{\,2}-q_2^{\,2})}{(4\pi^2+q_1^{\,2})(4\pi^2+q_2^{\,2})}\right]A^2\\ +\Bigg[\frac{4}{15\pi}(5\gamma_1+8\gamma_2+3\gamma_3)-4\pi\delta'(\gamma_1-\gamma_3)\left\{\frac{q_1\coth\frac{1}{2}q_1}{4\pi^2+q_2^{\,2}}-\frac{q_2\coth\frac{1}{2}q_2}{4\pi^2+q_1^{\,2}}\right\}\\ -32\pi^3\delta'\gamma_2\left\{\frac{q_2\coth\frac{1}{2}q_2}{(\pi^2+q_2^{\,2})(9\pi^2+q_2^{\,2})}-\frac{q_1\coth\frac{1}{2}q_1}{(\pi^2+q_1^{\,2})(9\pi^2+q_1^{\,2})}\right\}\Bigg]A+\Bigg[\tfrac{1}{2}(\gamma_1+\gamma_3)\\ -16\pi^2(\gamma_1-\gamma_3)\delta'\coth\tfrac{1}{2}q_1\coth\tfrac{1}{2}q_2\frac{a^2(q_1^{\,2}-q_2^{\,2})(10\pi^2+q_1^{\,2}+q_2^{\,2})}{(\pi^2+q_1^{\,2})(\pi^2+q_2^{\,2})(9\pi^2+q_1^{\,2})(9\pi^2+q_2^{\,2})}\Bigg].\end{aligned}$$
$$. \quad . \qquad (109)$$

As we have already stated R determined by formula (99) and eqns. (108) and (109) is 16 times the value of R for the case under consideration.

As in case (ii) the determination of R_c in accordance with eqns. (99), (108) and (109) was carried out in two stages: an approximation with no variational parameter and corresponding to a trial function

$$F=A \sin 2\pi\zeta, \qquad (110)$$

was first carried out. The formula for R in this first approximation (which can be obtained by letting $A\to\infty$ in eqns. (108) and (109)) is

$$R=\frac{(4\pi^2+a^2)[(4\pi^2+a^2)^2+4\pi^2Q]}{16a^2[1-16\pi^2\delta'(q_1{}^2-q_2{}^2)/(4\pi^2+q_1{}^2)(4\pi^2+q_2{}^2)]}. \qquad (111)$$

Table 3. Critical Rayleigh Numbers and Related Constants for the Case when One of the Bounding Surfaces is Free and the Other is Rigid

Q	a	A	First Approximation	Second Approximation	R_c/R_0
0	5·36	4·89	1112·7	1100·75	1·0000
10	5·50	4·96	1179·4	1167·2	1·0604
50	5·94	5·25	1428·3	1415·5	1·2859
100	6·34	5·59	1712·7	1699·4	1·5439
200	6·9	6·19	2231·3	2217·6	2·0146
500	8·0	7·54	3600·2	3586·1	3·2579
1000	9·0	9·21	5627·5	5613·3	5·0995
2000	10·2	11·6	9318·7	9304·5	8·4529
4000	11·5	14·9	16133·0	16119·0	14·644
6000	12·4	17·3	22606·0	22592·0	20·524
8000	13·0	19·4	28893·0	28879·0	26·236
10000	13·5	21·2	35058·0	35044·0	31·836
20000	15·3	27·9	64861·0	64847·0	58·912
40000	17·3	37·0	122155·0	122140·0	110·96

The value of a for which this formula gave a minimum value for R was first determined. Then with this value of a a second approximation was carried out with the aid of eqns. (108) and (109) by making the variation with respect to A. The results of such calculations are summarized in table 3.

It may be noted here that in the field-free case when $Q=0$ and $q_1=q_2=a$ the limiting form of (111), namely,

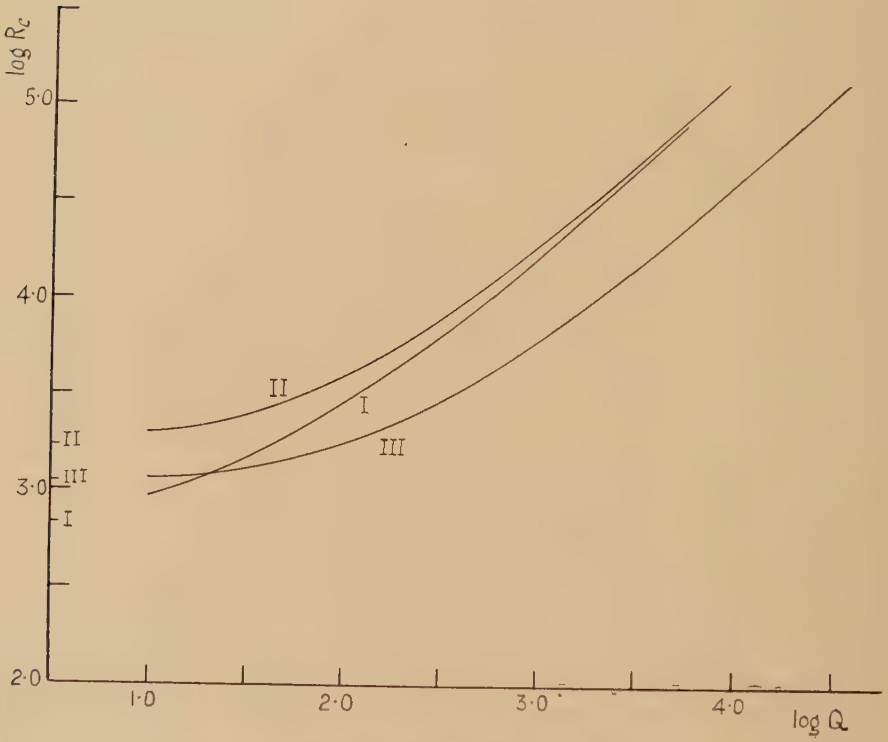
$$R=\frac{(4\pi^2+a^2)^3}{16a^2[1-64\pi^2a\sinh^2\frac{1}{2}a/(\sinh a-a)(4\pi^2+a^2)^2]} \qquad (112)$$

gives $R_c=1\,112{\cdot}7$ for $a=5{\cdot}36$; and this value should be compared with $R_c=1\,100{\cdot}65$ determined by Pellew and Southwell (1940) by an exact solution of the problem. On the other hand a second approximation carried out with the corresponding limiting forms of eqns. (108) and (109) gave the value 1100·75. As in case (ii) we may, therefore, expect that the tabulated values of R_c in the second approximation are correct to at least one part in 10^5.

Finally, it may be remarked that for sufficiently large Q the values of R_c for this case approach a quarter of that for case (ii) (or (i)).

The results of the calculations for all three cases are illustrated in fig. 1. It is apparent from this figure, as well as from tables 1 to 3, that the existence of a magnetic field will have, depending on the value of Q and the boundary conditions, an inhibiting effect on the onset of convection; and this inhibiting effect will be the greater the greater H and (or) σ are (is). In a general way the reason for this inhibiting effect is clear. When the field is strong (or the conductivity high) motions at right angles to **H** will become increasingly 'difficult' and this will prevent an 'easy' closing in of the stream lines required for convection. This

Fig. 1



The variation of the critical Rayleigh number R_c for the onset of instability as a function of Q for the three cases (i) both bounding surfaces free (curve marked I), (ii) both bounding surfaces rigid (curve marked II) and (iii) one bounding surface free and the other rigid (curve marked III). The points marked I, II and III on the R_c-axis are the limiting values for the three cases considered in the absence of a field.

also explains why when instability finally sets in, it does so for a value of a which is increasingly large: for, a large a means that the cells are elongated in the direction of **H** (cf. eqns. (46) and (49)) and motions in this direction are not hindered.

In order to see how effective the magnetic field will be in practical cases in inhibiting the onset of convection we shall consider mercury at room temperature. We have

$$\sigma=1{\cdot}1\times10^{-5},\quad \rho\nu=1{\cdot}7\times10^{-2}$$

and

$$Q=\frac{H^2\sigma d^2}{\rho\nu}=6\times10^{-4}\,H^2d^2, \qquad \text{(113)}$$

where it may be recalled that H is measured in gauss. For

$$d=1,\ H=10^3,\ Q=600; \qquad \text{(114)}$$

and the critical Rayleigh numbers for the three cases considered are

$$\left.\begin{aligned} R_c&=\ \ 9\,942,\ R_c/R_0=15{\cdot}12,\ \text{in case (i)},\\ R_c&=11\,660,\ R_c/R_0=\ \ 6{\cdot}80,\ \text{in case (ii)},\\ R_c&=\ \ 4\,018,\ R_c/R_0=\ \ 3{\cdot}65,\ \text{in case (iii)}.\end{aligned}\right\} \qquad \text{(115)}$$

It would appear from these values that the predicted effect should be easily detectable in the laboratory. In § 9 we shall indicate certain astrophysical and geophysical connections in which this effect may play a role.

§ 7. THE CASE WHEN H AND g ACT IN DIFFERENT DIRECTIONS

We shall now briefly consider the case when **H** and **g** act in different directions.

Let **H** act in a direction inclined at an angle ϑ to the vertical, i.e. to the direction in which **g** acts. Also, let the direction of the x-axis be so chosen that **H** lies in the xz-plane. Then

$$\boldsymbol{\lambda}=(0,0,1)\ \text{and}\ \boldsymbol{\tau}=(\sin\vartheta,0,\cos\vartheta), \qquad \text{(116)}$$

and eqns. (38) and (39) become

$$\nabla^4 w=-\frac{\gamma}{\nu}\left(\frac{\partial^2\theta}{\partial x^2}+\frac{\partial^2\theta}{\partial y^2}\right)+\frac{\mu^2H^2\sigma}{\rho_0\nu}\left(\cos\vartheta\frac{\partial}{\partial z}+\sin\vartheta\frac{\partial}{\partial x}\right)^2 w \qquad \text{(117)}$$

and

$$\nabla^6 w=\frac{|\beta|\gamma}{\kappa\nu}\left(\frac{\partial^2 w}{\partial x^2}+\frac{\partial^2 w}{\partial y^2}\right)+\frac{\mu^2H^2\sigma}{\rho_0\nu}\left(\cos\vartheta\frac{\partial}{\partial z}+\sin\vartheta\frac{\partial}{\partial x}\right)^2\nabla^2 w. \qquad \text{(118)}$$

If we seek solutions of eqns. (117) and (118) which are independent of x and are of the forms

$$w=W(z)\cos\frac{ay}{d}\ \text{and}\ \theta=\Theta(z)\cos\frac{ay}{d}, \qquad \text{(119)}$$

then eqns. (117) and (118) reduce to eqns. (50) and (51) with the only change that in the definition of Q (eqns. (52)) $H \cos \vartheta$ replaces H. Consequently, if we restrict ourselves to an onset of convection in the form of infinitely extended 'rolls' (instead of cells; in this connection see Chandra 1938) the discussion of the criterion for stability can be carried out exactly as in §§ 4 to 6; and the critical Rayleigh numbers can be deduced from those given in tables 1 to 3 by interpreting H to mean the component of $\mathbf{H}$ in the direction of $\mathbf{g}$. The question whether lower Rayleigh numbers can be reached by considering patterns of motion more complicated than rolls will be left open for the present though it would appear that on the whole the possibilities for this happening are rather remote.

§ 8. An Examination of the Principle of the Exchange of Stabilities

As we have already pointed out, the description of the situation in marginal stability in terms of the relevant equations in which $\partial/\partial t$ has been set equal to zero is valid only if the principle of the exchange of stabilities can be justified. Now this principle supposes that in the neighbourhood of the state in marginal stability the effects of a free disturbance of the steady state will vary like $e^{\omega t}$ where ω is a real number so that $\partial/\partial t = \omega = 0$ does define the marginal state. The principle will also be valid if it can be shown that even if ω is complex, the limit of the situations with a positive real part for ω is one in which both the real and the imaginary parts tend to zero simultaneously. But the principle will not be valid if when the real part of ω tends to zero from the positive side (i.e. $\Re(\omega) \to +0$) the imaginary part of ω tends to a finite limit. In this latter case when instability develops, the ensuing motions will be of the nature of oscillations with increasing amplitude, i.e. we shall have 'overstability' in the sense of Eddington.

For the classical Rayleigh–Jeffreys problem, Pellew and Southwell (1940) have proved by a very general argument that for all adverse temperature gradients (i.e. $\beta = -|\beta|$) the time dependent equations admit solutions varying like $e^{\omega t}$ with ω necessarily a real number. The application of the principle of the exchange of stabilities is therefore a valid procedure for that problem. On the other hand it will appear that the method of Pellew and Southwell is not quite strong enough to establish the principle for the problem on hand. We shall therefore examine a particular case in some detail and then see how far the general argument enables us to extend the deductions drawn from the particular case.

We start then from the general time dependent eqns. (28) to (30). Eliminating ϖ from the first of these equations by operating with ∇^2 and making use of eqn. (26), we obtain

$$\frac{\partial}{\partial t} \nabla^2 w = \frac{\mu H}{4\pi\rho_0} \tau_j \frac{\partial}{\partial x_j} \nabla^2 h + \gamma \left(\nabla^2 \theta - \lambda_i \lambda_j \frac{\partial^2 \theta}{\partial x_i \partial x_j} \right) + \nu \nabla^4 w. \quad . \quad . \quad (120)$$

If all quantities vary like $e^{\omega t}$ the foregoing equation and eqns. (29) and (30) reduce to

$$\left.\begin{aligned}(\omega-\nu\nabla^2)\nabla^2 w&=\frac{\mu H}{4\pi\rho_0}\tau_j\frac{\partial}{\partial x_j}\nabla^2 h+\gamma\left(\nabla^2\theta-\lambda_i\lambda_j\frac{\partial^2\theta}{\partial x_i\partial x_j}\right),\\(\omega-\kappa\nabla^2)\theta&=-\beta w,\\(\omega-\eta\nabla^2)h&=H\tau_j\frac{\partial w}{\partial x_j}.\end{aligned}\right\} \quad . \quad (121)$$

In the further discussion of this problem we shall restrict ourselves to the case when both **H** and **g** act in the direction of the vertical (0, 0, 1). Then eqns. (121) become

$$\left.\begin{aligned}(\omega-\nu\nabla^2)\nabla^2 w&=\frac{\mu H}{4\pi\rho_0}\frac{\partial}{\partial z}\nabla^2 h+\gamma\left(\frac{\partial^2\theta}{\partial x^2}+\frac{\partial^2\theta}{\partial y^2}\right),\\(\omega-\kappa\nabla^2)\theta&=-\beta w,\\(\omega-\eta\nabla^2)h&=H\frac{\partial w}{\partial z}.\end{aligned}\right\} \quad . \; . \; . \quad (122)$$

By operating the first of these equations by $(\omega-\kappa\nabla^2)(\omega-\eta\nabla^2)$ we can eliminate h and θ and obtain an equation only for w. Thus,

$$(\omega-\kappa\nabla^2)(\omega-\eta\nabla^2)(\omega-\nu\nabla^2)\nabla^2 w=\frac{\mu H^2}{4\pi\rho_0}\frac{\partial^2}{\partial z^2}(\omega-\kappa\nabla^2)\nabla^2 w$$
$$+|\beta|\gamma(\omega-\eta\nabla^2)\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}\right)w. \quad . \; . \; . \; . \; . \; . \quad (123)$$

We shall reduce this equation still further by assuming, as in § 4, that the variables x and y can be separated from z and that w is of the form

$$w=f(x,y)W(z) \quad . \; . \; . \; . \; . \; . \; . \; . \quad (124)$$

where f satisfies eqn. (44). For w of this form, eqn. (123) can be reduced to

$$[\Omega-\kappa(D^2-a^2)][\Omega-\eta(D^2-a^2)][\Omega-\nu(D^2-a^2)](D^2-a^2)W$$
$$=\frac{\mu H^2 d^2}{4\pi\rho_0}[\Omega-\kappa(D^2-a^2)]D^2(D^2-a^2)W-a^2 d^4|\beta|\gamma[\Omega-\eta(D^2-a^2)]W, \; . \quad (125)$$

where $\Omega=\omega d^2$, $z=\zeta d$ and $D=\partial/\partial\zeta$. . . . . . . (126)

A solution of eqn. (125) must be sought which satisfies the various boundary conditions of the problem. By letting

$$W=\text{constant}\sin\pi\zeta \quad . \; . \; . \; . \; . \; . \; . \; . \quad (127)$$

we can satisfy the boundary conditions appropriate to both bounding surfaces being free (i.e. $\theta=W=D^2W=0$ for $\zeta=0$ and 1) and further have the normal component of the gradient of h vanish on the bounding surfaces. A straightforward discussion of eqn. (125) for any other set of boundary conditions is likely to be troublesome. It will not be attempted in this paper.

For W given by (127) eqn. (125) becomes

$$[\Omega+\kappa(\pi^2+a^2)]\,[\Omega+\eta(\pi^2+a^2)]\,[\Omega+\nu(\pi^2+a^2)]$$
$$+\frac{\mu H^2 d^2}{4\pi\rho_0}\pi^2[\Omega+\kappa(\pi^2+a^2)]-|\beta|\gamma\, d^4\frac{a^2}{a^2+\pi^2}[\Omega+\eta(\pi^2+a^2)]=0. \quad (128)$$

Letting

$$q=\frac{\Omega}{\pi^2}=\omega\frac{d^2}{\pi^2} \text{ and } y=\frac{a^2}{\pi^2}+1, \quad \ldots\ldots \quad (129)$$

we can rewrite eqn. (128) in the form

$$(q+\kappa y)(q+\eta y)(q+\nu y)+Q_1\eta\nu(q+\kappa y)$$
$$-R_1\kappa\nu(q+\eta y)\frac{y-1}{y}=0, \quad \ldots\ldots \quad (130)$$

where (cf. eqns. (52))

$$Q_1=\frac{Q}{\pi^2}=\mu^2\frac{H^2\sigma\, d^2}{\rho_0\nu\,\pi^2} \text{ and } R_1=\frac{R}{\pi^4}=\frac{|\beta|\gamma\, d^4}{\kappa\nu\,\pi^4}. \quad \ldots \quad (131)$$

Equation (130) is a cubic equation for q. Writing it in the form

$$q^3+Bq^2+Cq+D=0, \quad \ldots\ldots \quad (132)$$

we have

$$\left.\begin{aligned} D&=\kappa\nu\eta\{y^3+Q_1y-R_1(y-1)\},\\ C&=y^2s_2+Q_1\eta\nu-R_1\kappa\nu\frac{y-1}{y},\\ B&=ys_1,\ s_1=\kappa+\nu+\eta \text{ and } s_2=\kappa\nu+\nu\eta+\eta\kappa. \end{aligned}\right\} \quad \ldots \quad (133)$$

From eqn. (132) it follows that if $D<0$, it admits a positive real root. Consequently, if $D<0$ and we let it tend to zero from the negative side we shall have $q\to 0$ through positive real values. In other words, in this case we shall approach marginal stability as the limit of solutions which diverge as $e^{\omega t}$ with ω real and positive; $D<0$ is therefore a sufficient condition for the principle of the exchange of stabilities to be valid. Therefore, let

$$D>0. \quad \ldots\ldots \quad (134)$$

Equation (132) then allows a negative real root and the corresponding solutions lead to an exponential damping of an initial disturbance; these solutions (as well as their limit $D\to+0$) are of no interest to us. We have now to examine whether under these conditions the complex roots of (132) (if such exist) can have a positive real part. (For a similar discussion of a cubic equation see Jeans (1929, p. 120).) If $p/2$ denotes the real part of a complex root of eqn. (132), then it satisfies the equation

$$p^3+2Bp^2+(B^2+C)p+(BC-D)=0. \quad \ldots \quad (135)$$

If now

$$BC-D<0, \quad \ldots\ldots \quad (136)$$

eqn. (135) will admit a positive real root and by letting $BC-D\to 0$ through negative values, we shall have the case when a complex root of eqn. (132), having a positive real part, tends to a purely imaginary limit. Thus, if

$$D>0 \text{ and } BC-D<0, \qquad . \quad . \quad . \quad . \quad . \quad . \quad (137)$$

the principle of the exchange of stabilities cannot be applied and we will have a case of over-stability.

According to eqns. (133), the conditions (137) are equivalent to

$$R_1(y-1)<y^3+Q_1y \qquad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (138)$$

and

$$\kappa\nu\eta\{y^3+Q_1y-R_1(y-1)\}>y^3s_1s_2+Q_1\eta\nu s_1y-R_1\kappa\nu s_1(y-1). \quad . \quad (139)$$

Rearranging this last inequality, we have

$$R_1(y-1)>\frac{1}{\kappa\nu(\kappa+\nu)}\{y^3(s_1s_2-\kappa\nu\eta)+Q_1\eta\nu(\eta+\nu)y\}. \quad . \quad . \quad (140)$$

The inequalities (138) and (140) can now be combined in the form

$$y^3+Q_1y>R_1(y-1)>\frac{1}{\kappa\nu(\kappa+\nu)}\{y^3(s_1s_2-\kappa\nu\eta)+Q_1\eta\nu(\eta+\nu)y\}. \quad (141)$$

Since y is necessarily positive (cf. eqn. (129)), we can rearrange the outer inequalities of (141) to give

$$Q_1\left\{1-\frac{\eta(\eta+\nu)}{\kappa(\kappa+\nu)}\right\}>y^2\frac{\eta}{\kappa\nu(\kappa+\nu)}[\nu(\eta+\nu)+\kappa(\kappa+\eta)+2\kappa\nu]. \quad . \quad (142)$$

This last inequality is a necessary condition for instability to arise through oscillations of increasing amplitude; it is therefore a condition for the principle of the exchange of stabilities *not* to be valid.

According to (142) a sufficient condition for the principle of the exchange of stabilities to be valid is

$$\frac{\eta(\eta+\nu)}{\kappa(\kappa+\nu)}>1 \text{ or } \eta>\kappa. \qquad . \quad . \quad . \quad . \quad . \quad . \quad (143)$$

(In § 10 we shall show that this condition is also necessary.) It would appear that for fluids under terrestial conditions this last inequality, $\eta>\kappa$, will be satisfied with a large factor of safety. Thus, for mercury at room temperature

$$\eta=7{\cdot}5\times 10^3 \text{ cm}^2/\text{sec and } \kappa=4{\cdot}7\times 10^{-2} \text{ cm}^2/\text{sec}. \quad . \quad . \quad (144)$$

Accordingly, in these cases we may safely apply the principle of the exchange of stabilities. However, under astrophysical conditions (see § 9) $\eta\ll\kappa$ and we must go back to the general inequality (142); we shall return to this in § 9.

If condition (142) should be met and we have a case of over-stability, then the frequency of oscillation ω' at marginal stability will be given by

$$\omega'=\frac{\pi^2}{d^2}\left\{\frac{\kappa\nu\eta}{ys_1}[y^3+Q_1y-R_1(y-1)]\right\}^{1/2}$$

$$=\frac{\pi^2}{d^2}\left\{y^2s_2+Q_1\eta\nu-R_1\kappa\nu\frac{y-1}{y}\right\}^{1/2}, \quad . \quad . \quad . \quad (145)$$

and the critical Rayleigh number for the onset of convection can be obtained by minimizing (with respect to y) the expression

$$R=\pi^4\frac{s_1s_2-\kappa\nu\eta}{\kappa\nu(\kappa+\nu)}\,\frac{y^3+Q'y}{y-1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (146)$$

where

$$Q'=\frac{Q}{\pi^2}\,\frac{\eta\nu(\eta+\nu)}{s_1s_2-\kappa\nu\eta}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (147)$$

The foregoing discussion applies to solutions of eqns. (122) which satisfy only a particular set of boundary conditions. For solutions satisfying more general boundary conditions we might treat eqns. (122) in the manner of Pellew and Southwell (1940, §§ 16 to 18). Thus if ω is complex, let w^*, θ^* and h^* belong to the complex conjugate, ω^*, of ω. Then (cf. eqns. (122))

$$(\omega^*-\nu\nabla^2)\nabla^2w^*-\frac{\mu H}{4\pi\rho_0}\frac{\partial}{\partial z}\nabla^2h^*=\gamma\left(\frac{\partial^2\theta^*}{\partial x^2}+\frac{\partial^2\theta^*}{\partial y^2}\right). \quad . \quad . \quad (148)$$

Multiplying this equation crosswise by the second of the eqns. (122) we obtain

$$\beta w(\omega^*-\nu\nabla^2)\nabla^2w^*-\frac{\mu H\beta}{4\pi\rho_0}w\frac{\partial}{\partial z}\nabla^2h^*+\gamma(\omega-\kappa\nabla^2)\theta\left(\frac{\partial^2\theta^*}{\partial x^2}+\frac{\partial^2\theta^*}{\partial y^2}\right)=0. \quad (149)$$

Now we integrate this equation over a unit cell. The various volume integrals can be reduced by transforming them in accordance with Green's theorem and making use of the proper boundary conditions on the cell walls and the bounding surfaces. The integrals arising from the first and the last terms on the left-hand side of (149) have already been reduced by Pellew and Southwell. It remains to consider

$$\iiint w\frac{\partial}{\partial z}\nabla^2h^*\,dx\,dy\,dz. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (150)$$

Integrating this by parts (with respect to z) and remembering that w must in all cases vanish on the bounding surfaces, we have

$$-\iiint\frac{\partial w}{\partial z}\nabla^2h^*\,dx\,dy\,dz. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (151)$$

Replacing $\partial w/\partial z$ by $(\omega-\eta\nabla^2)h/H$ (cf. the last of the eqns. (122)) and transforming the integral by Green's theorem, we obtain

$$\frac{1}{H}\iiint\{\omega\,|\,\mathrm{grad}\,h\,|^2+\eta\,|\,\nabla^2h\,|^2\}\,dx\,dy\,dz-\frac{\omega}{H}\iint h\,\mathrm{grad}\,h^*\;d\mathbf{S}. \quad (152)$$

The surface integral in (152) vanishes since on the cell walls (being surfaces of symmetry) $\partial h^*/\partial n$ vanishes and on the bounding surfaces ($\zeta=0$ and 1) either h or $\partial h/\partial z$ must vanish. We are therefore left with

$$\iiint w\frac{\partial}{\partial z}\nabla^2 h^*\,dx\,dy\,dz=\frac{1}{H}(\omega I_5+\eta I_6), \quad . \quad . \quad . \quad . \quad . \quad (153)$$

where I_5 and I_6 denote the positive definite integrals

$$I_5=\iiint|\operatorname{grad} h|^2\,dx\,dy\,dz \text{ and } I_6=\iiint|\nabla^2 h|^2\,dx\,dy\,dz. \quad (154)$$

The result of integrating eqn. (149) over a unit cell is therefore (cf. Pellew and Southwell 1940, eqn. (29))

$$\beta\omega^* I_1+\frac{\mu\beta\omega}{4\pi\rho_0}I_5+\gamma\omega I_3+\beta\nu I_2+\kappa\gamma I_4+\frac{\mu\beta\eta}{4\pi\rho_0}I_6=0, \quad . \quad . \quad (155)$$

where

$$\left.\begin{aligned} I_1&=\iiint|\operatorname{grad} w|^2\,dx\,dy\,dz, \quad I_2=\iiint|\nabla^2 w|^2\,dx\,dy\,dz,\\ I_3&=\frac{a^2}{d^2}\iiint|\theta|^2\,dx\,dy\,dz \text{ and } I_4=\frac{a^2}{d^2}\iiint|\operatorname{grad}\theta|^2\,dx\,dy\,dz\end{aligned}\right\} \quad (156)$$

are all positive definite integrals.

Separating the real and the imaginary parts of eqn. (155), we have

$$\mathfrak{R}(\omega)\left\{\gamma I_3+\beta\left(I_1+\frac{\mu I_5}{4\pi\rho_0}\right)\right\}+\beta\left(\nu I_2+\frac{\mu\eta I_6}{4\pi\rho_0}\right)+\kappa\gamma I_4=0 \quad . \quad (157)$$

and

$$\mathfrak{J}(\omega)\left\{-\beta\left(I_1-\frac{\mu I_5}{4\pi\rho_0}\right)+\gamma I_3\right\}=0. \quad . \quad . \quad . \quad (158)$$

From eqn. (158) it follows that $\mathfrak{R}(\omega)$ is necessarily negative if $\beta\geqslant 0$; in other words when the temperature gradient is not adverse all small disturbances will decay. For $\mathfrak{R}(\omega)$ to be positive it is a necessary condition that β be negative. Let $\beta=-|\beta|$. Then eqns. (157) and (158) give

$$\mathfrak{R}(\omega)=\frac{|\beta|(\nu I_2+\mu\eta I_6/4\pi\rho_0)-\kappa\gamma I_4}{\gamma I_3-|\beta|(I_1+\mu I_5/4\pi\rho_0)} \quad . \quad . \quad . \quad . \quad (159)$$

and

$$\mathfrak{J}(\omega)\left[|\beta|\left(I_1-\frac{\mu I_5}{4\pi\rho_0}\right)+\gamma I_3\right]=0. \quad . \quad . \quad . \quad (160)$$

From eqn. (160) we conclude that

either $$\mathfrak{J}(\omega)=0 \text{ or } |\beta|\left(I_1-\frac{\mu I_5}{4\pi\rho_0}\right)+\gamma I_3=0. \quad . \quad . \quad . \quad (161)$$

In the former case we recover the principle of the exchange of stabilities and we may readily verify that eqn. (159) then reduces to the formula (69) for evaluating R in the case of marginal stability. In the latter case, we may have over-stability; and a necessary condition for this to be possible is

$$\frac{\mu}{4\pi\rho_0}I_5>I_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (162)$$

or, explicitly (cf. eqns. (154) and (156))

$$\frac{\mu}{4\pi\rho_0}\iiint |\operatorname{grad} h|^2\, dx\, dy\, dz > \iiint |\operatorname{grad} w|^2\, dx\, dy\, dz. \quad . \quad (163)$$

It would appear that this last condition can in general be satisfied only if the magnetic energy associated with the disturbance exceeds the kinetic energy. Under most conditions we should not expect this to happen: for, even in a magneto-hydrodynamic wave the two energies are only equal. We conclude that it is highly probable that the principle of the exchange of stabilities will be valid under most terrestial conditions; but it is by no means certain that this will be the case under astrophysical conditions. A further examination of the inequality (142) in § 9 in the light of (163) confirms this conclusion.

§9. The Inhibition of Convection by a Magnetic Field under Astrophysical Conditions

On the grounds that motions ensuing on the onset of convection must be characterized by closed stream lines and that such motions would be hindered by a magnetic field, it has been suggested (cf. Biermann 1941 and Walén 1949*) that the presence of a magnetic field might suppress convection altogether. Indeed, one has sought to explain the appearance of sunspots in terms of such an inhibiting effect of the magnetic fields associated with them. The theory developed in the preceding sections enables us to examine to what extent these suggestions are sound.

Consider, then, the conditions in the so-called hydrogen-convection zone in the solar atmosphere where

$$\rho \simeq 2\times 10^{-7}\ \text{g/cm}^3,\ T=6\,000^\circ\ \text{K} \quad . \quad . \quad . \quad . \quad (164)$$

and the most abundant constituent of the atmosphere, namely, hydrogen is getting about half ionized. Osterbrock† (1952) estimates that under these conditions

$$\kappa \simeq 10^{13}\ \text{cm}^2/\text{sec},\ \nu \simeq 5\times 10^3\ \text{cm}^2/\text{sec and } \eta \simeq 7\times 10^6\ \text{cm}^2\,\text{sec}. \quad . \quad (165)$$

It will be noticed that contrary to terrestial conditions (cf. eqns. (113) and (144)) the largest of these coefficients, by far, is the thermometric conductivity, κ. The reason for this is that under astrophysical conditions the transport of heat by radiation is an extremely efficient process. Indeed, it would appear that

$$\kappa \gg \eta \gg \nu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (166)$$

is characteristic of astrophysical conditions. We shall now examine in some detail the extent and manner of the inhibitory influence of magnetic fields on convection when (166) holds.

---

* I am indebted to Professor T. G. Cowling for drawing my attention to these references.

† I am grateful to Mr. D. Osterbrock for allowing me to use the results of his investigation before publication.

The first question to settle is of course the applicability of the principle of the exchange of stabilities. For this purpose we go back to the fundamental inequality (142). It will be recalled that this inequality was derived for boundary conditions appropriate to both bounding surfaces being free; it would not seem that the use of these boundary conditions for the astrophysical problem is artificial.

Now when (166) holds, the inequality (142) can be simplified to

$$Q_1>y^2\frac{\eta}{\nu}. \tag{167}$$

Suppose that this condition is satisfied and that over-stability arises. As we have already pointed out, under these circumstances the critical Rayleigh number at which instability will first develop can be found by minimizing the expression on the right-hand side of eqn. (146). And since $(s_1s_2-\kappa\nu\eta)\simeq\kappa^2\eta$, when (166) holds, we can simplify eqns. (146) and (147) to the forms (cf. eqns. (131))

$$R_1=\frac{\eta}{\nu}\frac{y^3+Q'y}{y-1} \quad \text{where } Q'=Q_1\frac{\eta\nu}{\kappa^2}. \tag{168}$$

For Q_1 a multiple of η/ν say $A\eta/\nu$, $Q'=A\eta^2/\kappa^2$; we can, therefore, neglect Q' in the expression for R_1 so long as A does not become comparable to κ^2/η^2. Consequently, for a considerable range of Q_1 (satisfying (167)) we can write

$$R_1=\frac{\eta}{\nu}\frac{y^3}{y-1} \quad \left(Q_1<\frac{\kappa^2}{\eta\nu}\right). \tag{169}$$

The right-hand side of this expression has a minimum for $y=3/2$. Therefore, under the conditions specified, the critical Rayleigh number for the onset of over-stability is

$$R_c=\pi^4R_{1,\,\text{min.}}=\frac{27}{4}\pi^4\frac{\eta}{\nu}; \tag{170}$$

and this is independent of the strength of the prevailing magnetic field.

Returning to the more general inequality (141), we observe that under marginal conditions this must reduce to

$$y^3+Q_1y>R_1(y-1)=\frac{1}{\kappa\nu(\kappa+\nu)}\{y^3(s_1s_2-\kappa\nu\eta)+Q_1\eta\nu(\eta+\nu)y\}. \tag{171}$$

We have seen that when (166) holds and $Q_1<\kappa^2/\eta\nu$ marginal stability obtains for $y=3/2$ and the term in Q_1 on the right-hand side of the equality in (171) can be neglected. Consequently, for

$$1\ll\eta/\nu<Q_1<\kappa^2/\eta\nu, \tag{172}$$

(171) becomes

$$\tfrac{3}{2}Q_1>\tfrac{1}{2}R_{1,\,\text{min.}}=\frac{27}{8}\frac{\eta}{\nu}; \tag{173}$$

and this is consistent with (170).

What the discussion in the preceding paragraph establishes is this: when (166) holds, instability in the form of oscillations of increasing amplitude *can* arise for

$$Q_1 > \frac{9\eta}{4\nu}; \qquad (174)$$

and when it does develop by over-stability, the critical Rayleigh number remains practically constant and has the value

$$R_c = \frac{27}{4}\pi^4\frac{\eta}{\nu} = 657{\cdot}5\frac{\eta}{\nu}, \qquad (175)$$

so long as

$$Q_1 < \kappa^2/\eta\nu. \qquad (176)$$

For $Q_1 > \kappa^2/\eta\nu$, R_c increases in a manner which is identical with case (i) treated in § 6; the results of table 1 can in fact be used quite generally if Q is interpreted as meaning $Q'\pi^2$ and R_c as the required Rayleigh number divided by $(s_1 s_2 - \kappa\nu\eta)/\kappa\nu(\kappa+\nu)$.

On the other hand if $Q_1 < 9\eta/4\nu$, the principle of the exchange of stabilities will be applicable and instability at the critical Rayleigh number will arise through ordinary cellular convection in accordance with the results of § 6. Again, restricting ourselves to the case when both the bounding surfaces are free and the results for case (i) § 6 apply, we can, in the relevant range of $Q_1(\sim\eta/\nu \gg 1)$, write (cf. eqn. (87))

$$R_c = \pi^4 Q_1. \qquad (177)$$

The value of R_c given by (177) is less than that given by (175) so long as

$$Q_1 \leqslant Q_1{}^* = \frac{27}{4}\frac{\eta}{\nu}. \qquad (178)$$

For $Q_1 > Q_1{}^*$ the reverse will be the case. We may therefore conclude that for $Q_1 \leqslant Q_1{}^*$ instability will arise at the appropriate Rayleigh number by cellular convection while for $Q_1 > Q_1{}^*$ it will arise through over-stability.

We have seen that when the conditions (166) hold, the critical Rayleigh number for instability remains constant for a considerable range of Q_1 greater than $Q_1{}^*$. But Q_1 (i.e. the magnetic field) influences the frequency of the oscillations with which over-stability sets in. Thus, according to eqn. (145), for $Q_1 \gg Q_1{}^*$

$$\omega' \simeq \frac{\pi^2}{d^2}(Q_1\eta\nu)^{1/2}. \qquad (179)$$

The meaning of this frequency becomes clearer when we express Q_1 in terms of the velocity of the magneto-hydrodynamic wave, namely

$$V_m = \left(\frac{\mu H^2}{4\pi\rho_0}\right)^{1/2}. \qquad (180)$$

Thus (cf. eqns. (20) and (131))

$$Q_1 = \frac{\mu H^2}{4\pi\rho_0\eta\nu}\frac{d^2}{\pi^2} = \frac{V_m{}^2}{\eta\nu}\frac{d^2}{\pi^2}. \qquad (181)$$

Substituting this last expression for Q_1 in (179) we find,

$$\omega' = \frac{\pi}{d} V_m ; \qquad \dots\dots\dots\dots \quad (182)$$

$2\pi/\omega'$ is therefore the time required for a magneto-hydrodynamic wave to travel a distance equal to $\frac{1}{2}d$. Similarly, some insight into the meaning of Q_1^* (at which instability by convection changes into instability by oscillations of increasing amplitude) can be obtained by expressing it in terms of the velocity of the magneto-hydrodynamic wave. Thus, letting

$$Q_1^* = \frac{27}{4}\frac{\eta}{\nu} = \frac{(V_m^*)^2}{\eta\nu}\frac{d^2}{\pi^2}, \qquad \dots\dots\dots \quad (183)$$

we have

$$V_m^* = \frac{\sqrt{27}}{2}\frac{\pi}{d}\eta . \qquad \dots\dots\dots\dots \quad (184)$$

We may now summarize the results of the foregoing discussion as follows: so long as the velocity of the magneto-hydrodynamic wave is less than V_m^* given by (184), instability will arise through cellular convection when the appropriate Rayleigh number derived on the principle of the exchange of stabilities is reached. But if $V_m^* < V_m < \pi\kappa/d$ we will have over-stability when the Rayleigh number reaches the value $27\pi^4\eta/4\nu$; and the frequency of oscillation at marginal stability will be essentially determined by the time required for the magneto-hydrodynamic wave to travel the thickness of the atmosphere. All these conclusions, it should be repeated, are valid only if the (166) holds.

Returning to the numerical values (165) for the basic coefficients, we observe that

$$R_c \sim 10^6 \text{ for } 3\times 10^{16} < Q < 10^5. \qquad \dots\dots\dots \quad (185)$$

Therefore, under astrophysical conditions, the inhibition of the onset of *instability* by a magnetic field is *not* very pronounced. On the other hand since in all practical cases we may expect the instability to develop in the form of oscillations of increasing amplitude, the question whether in these cases we will have convection in the true sense is not certain. The possibility of explaining sunspots in the manner indicated at the beginning of this section must therefore be left open for the present.

§10. THE MANNER OF THE ONSET OF INSTABILITY IN THE GENERAL CASE

In §§ 8 and 9 we have seen how in virtue of the inequalities $\eta > \kappa$ and $\kappa \gg \eta \gg \nu$, which are characteristic, respectively, of terrestial and astrophysical conditions, we are able to conclude that while in the former case instability always arises as cellular convection, in the latter case it will develop either as cellular convection or as over-stability depending on the magnitude of the velocity of the magneto-hydrodynamic wave. We shall now show how under quite general conditions we can distinguish the manner of the onset of instability.

As in §§ 8 and 9 we shall again restrict ourselves to the case when both surfaces confining the fluid are free. As we have seen, if instability should develop as cellular convection, then the critical Rayleigh number for instability is determined by the minimum of the right-hand side of the expression (cf. eqn. (84))

$$R^{(\mathrm{con})}=\pi^4\frac{y}{y-1}\left(y^2+\frac{Q}{\pi^2}\right), \quad . \quad . \quad . \quad . \quad . \quad . \quad (186)$$

where for comparison with (141) we have replaced $(1+x)$ in eqn. (84) by y. On the other hand, if instability should develop as over-stability then the critical Rayleigh number for instability is determined by the minimum of the right-hand side of the expression (cf. eqn. 146))

$$R^{(0-s)}=\pi^4\frac{s_1s_2-\kappa\nu\eta}{\kappa\nu(\kappa+\nu)}\frac{y}{y-1}\left\{y^2+\frac{Q}{\pi^2}\frac{\eta\nu(\eta+\nu)}{s_1s_2-\kappa\nu\eta}\right\}. \quad . \quad . \quad . \quad (187)$$

In eqns. (186) and (187) we have attached superscripts (con) and (0—s) to R to emphasize the nature of the instability which follows.

Using the results of table 1 we can readily draw in the (R_c, Q)-plane the critical Rayleigh numbers derived on the basis of either of the eqns. (186) and (187). Let the curves so derived be called the ‘ convection curve ’ and the ‘ over-stability curve ’ respectively. Both are monotonically increasing curves. The convection curve starts on the R_c-axis at the point

$$R_c^{(\mathrm{con})}=\frac{27}{4}\pi^4 \qquad (Q=0), \quad . \quad . \quad . \quad (188)$$

and for $Q\rightarrow\infty$ becomes asymptotic to the line

$$R_c^{(\mathrm{con})}=\pi^2Q \qquad (Q\rightarrow\infty). \quad . \quad . \quad . \quad (189)$$

On the other hand, the over-stability curve starts on the R_c-axis at

$$R_c^{(0-s)}=\frac{27}{4}\pi^4\frac{s_1s_2-\kappa\nu\eta}{\kappa\nu(\kappa+\nu)} \qquad (Q=0), \quad . \quad . \quad . \quad (190)$$

and for $Q\rightarrow\infty$ becomes asymptotic to the line

$$R_c^{(0-s)}=\pi^2\frac{\eta(\eta+\nu)}{\kappa(\kappa+\nu)}Q \qquad (Q\rightarrow\infty). \quad . \quad . \quad . \quad (191)$$

Since

$$\frac{s_1s_2-\kappa\nu\eta}{\kappa\nu(\kappa+\nu)}=1+\frac{\eta}{\kappa\nu(\kappa+\nu)}\{\nu(\eta+\nu)+\kappa(\kappa+\eta)+2\kappa\nu\}>1, \quad . \quad . \quad . \quad (192)$$

it follows that the over-stability curve *always* starts above the convection curve. Consequently, we need distinguish between only two cases :

(i) *The over-stability curve lies entirely above the convection curve.*—From eqns. (189) and (191) it follows that this will be the case whenever

$$\eta\geqslant\kappa. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (193)$$

Moreover, it is clear that in this case, when instability arises it will always be as cellular convection. This is what happens under normal terrestial conditions.

(ii) *The over-stability curve intersects the convection curve and for $Q \to \infty$ lies below it.*—According to eqns. (189) and (190) this will be the case whenever

$$\eta < \kappa. \qquad (194)$$

Conversely, whenever $\eta < \kappa$ the two curves must intersect. Let the value of Q at the point where the two curves intersect be Q^*. Then it is clear that for $Q \leqslant Q^*$, instability, when it arises, will be as cellular convection; but for $Q > Q^*$ we shall have over-stability when the critical Rayleigh number $R_c^{(0-s)}$ is reached. This is essentially what happens (though in a somewhat 'extreme' form) under astrophysical conditions.

§11. CONCLUDING REMARKS

From the numerical example considered in § 6 (eqns. (113) to (115)) it would appear that under normal terrestial conditions ($\eta > \kappa$) the inhibition of convection by a magnetic field is a very pronounced effect. And since convection and turbulence are often postulated in the earth's core (cf. Bullard 1948) it is possible that a study of the problem considered in this paper for a spherical geometry may lead to results of some significance. The problem of the instability of a fluid sphere heated within has recently been investigated by Jeffreys and Bland (1951); and this investigation provides an excellent context in which the theory developed in this paper can be extended. The consideration of this same problem when $\kappa > \eta$ and over-stability is possible may also be of interest for astrophysics. However, under astrophysical conditions ($\kappa \gg \eta \gg \nu$) the effects considered in this paper are probably of most direct application to sunspot phenomena. In particular, a re-examination of the problem of thermal stability of a fluid enclosed in a long circular cylinder which has been considered by Hales (1937) in connection with a theory of geysers may be instructive. A problem of a different kind relates to the nature of turbulence which will finally establish itself in an extended atmosphere on account of thermal instability. A beginning in this direction has recently been made by the writer (Chandrasekhar 1952). An extension of this investigation to the case when a magnetic field is present would be of interest. And finally, one might also investigate the related problem (cf. Low 1929) of the stability of a liquid between two co-axial cylinders rotating at different rates, when the liquid considered is an electrical conductor and a magnetic field is present; as is well known this problem in the absence of a field has been investigated by Taylor (1923). The author hopes to publish the results of his investigation of these various problems suggested by this paper in the near future.

In concluding this paper, I wish to record my indebtedness to Miss Donna Elbert for her very valuable assistance in having carried out all the numerical calculations involved in the preparation of tables 1 to 3.

References

Bénard, M., 1900, *Rev. gén. Sci. por. appl.*, **12,** 1261, 1309; 1901, *Ann. Chim. (Phys.)*, **23,** 62.

Biermann, L., 1941, *Vierteljahrsschrift d. Astron. Gesellschaft*, **76,** 194.

Boussinesq, J., 1903, *Theórie Analytique de la Chaleur*, **2,** 172. Paris.

Bullard, E. C., 1948, *Mon. Not. R. Astr. Soc. Geophys. Suppl.*, **5,** 248.

Chandra, K., 1938, *Proc. Roy. Soc.* A, **164,** 231.

Chandrasekhar, S., 1952, *Phil. Trans.* A, (in the press).

Christopherson, D. G., 1940, *Quart. J. Math.*, **11,** 63.

Hales, A. L., 1937, *Mon. Not. R. Astr. Soc. Geophys. Suppl.*, **4,** 122.

Jeans, J. H., 1929, *Astronomy and Cosmogony.* Cambridge University Press.

Jeffreys, H., 1926, *Phil. Mag.*, [7], **2,** 833; 1928, *Proc. Roy. Soc.* A. **118,** 195; 1930, *Proc. Camb. Phil. Soc.*, **26,** 170.

Jeffreys, H., and Bland, M. E. M., 1951, *Mon. Not. R. Astr. Soc. Geophys. Suppl.*, **6,** 148.

Low, A. R., 1929, *Proc. Roy. Soc.* A, **125,** 180.

Osterbrock, D. E., 1952, *Phys. Rev.* (in the press).

Pellew, A., and Southwell, R. V., 1940, *Proc. Roy. Soc.* A, **176,** 312.

Rayleigh, Lord, 1916, *Phil. Mag.*, [6], **32,** 529; also *Scientific Papers*, **6,** 432.

Schmidt, R. J., and Milverton, S. W., 1935, *Proc. Roy. Soc.* A, **152,** 586.

Sutton, O. G., 1950, *Proc. Roy. Soc.* A, **204,** 297.

Taylor, G. I., 1923, *Phil. Trans.* A, **223,** 289.

Thompson, W. B., 1951, *Phil. Mag.*, [7], **42,** 1417.

Walén, C., 1949, *On the Vibratory Rotation of the Sun.* Henrik Lindståhls Bokhandel. Stockholm.

XLVII. *Grain Density in Tracks of Electrons Moving at Relativistic Velocities in Photographic Emulsions*

By A. H. MORRISH *
H. H. Wills Physical Laboratory, University of Bristol †

[Received February 21, 1952]

SUMMARY

The variation in the grain-density of tracks produced in photographic plates by electrons movin gat relativistic velocities has been determined. It is found that the grain density increases by about 5% between the energies 9 and 25, measured in rest mass units, thereafter remaining constant at least up to $E/mc^2 \sim 10^4$.

§ 1. INTRODUCTION

PARTICLES of charge $/e/$ moving in photographic emulsions with velocities close to that of light, produce tracks in which the grain-density is close to the minimum value. The precise variation of the grain-density as a function of energy is of interest because it may provide a method of mass discrimination in certain energy regions, and because it is a problem which lends itself to quantitative theoretical calculation. Electron tracks are the most suitable for an investigation extending over a wide range of values of E/mc^2.

Ilford G5 emulsions, 400 microns thick were exposed at an altitude of 100 000 feet. After suitable processing *one* 6 in. × 6 in. plate was cut into four pieces, 3 in. × 3 in. The methods described by Hooper *et al.* (1951), were used in a systematic scan of certain regions, indicated by the cross-hatched areas of fig. 1, for pairs of electrons produced by the conversion of energetic photons. The close pairs of tracks thus produced can be attributed to electrons with great confidence. The energy of the pairs of electrons was determined by the multiple-scattering technique (Gottstein *et al.* 1951). In this way, more than 300 electron tracks, each at least 2 mm long and with values of E/mc^2 between 20 and $1{\cdot}5 \times 10^4$, were examined. Electron tracks for which $12 > E/mc^2 > 6$ were obtained by measuring single, highly scattered tracks with an ionization close to the minimum value, found in the same region of the plates.

§ 2. PLATE CALIBRATION

Since the effect to be investigated is a small one, considerable time was spent in determining the variation of grain-density with depth and position in the emulsion. For this purpose, electron tracks in the region $E/mc^2 \sim 500$ to 1 500 were used, the specific ionization being assumed constant in this range.

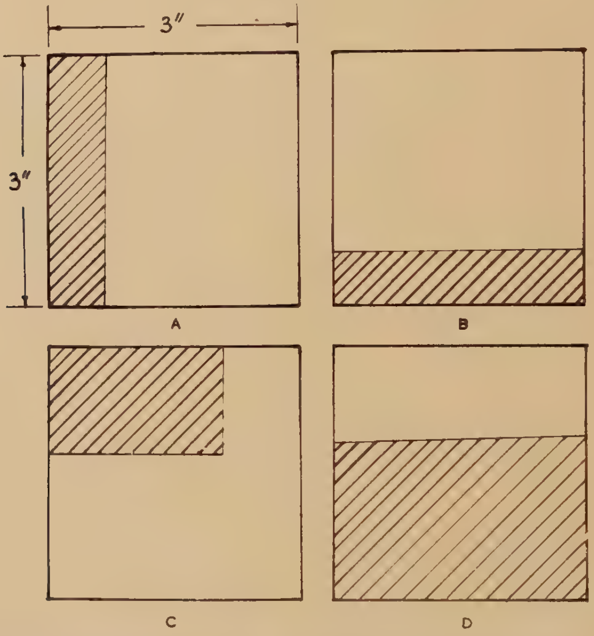
---

* Now at the University of British Columbia, Canada.
† Communicated by Professor C. F. Powell, F.R.S.

A grain was counted as one unit irrespective of size. This convention, useful up to 1·5× the minimum value, has the following advantages :—

(1) It allows rapid counting, no time being lost in estimating the number of grains in aggregates. (2) Subjective errors are small. The same individual tracks, recounted weeks later, always agreed with the earlier count to within 0·5%. (3) The variation of the grain-density with depth is much smaller than when a convention which 'weights' various grain-sizes is used. This is presumably associated with the fact that grains near the surface of the emulsion are commonly larger than those near the glass surface, because of greater physical development.

Fig. 1



The cross-hatched areas indicate the regions scanned for electron tracks. The four 3 in. × 3 in. plates were originally part of a single plate.

The depth variation was investigated for the cross-hatched region of plate *D* in fig. 1. The top 50 and bottom 50 microns of the emulsion were not included in this calibration. The remaining 300 microns were considered as made up of three 100 micron layers. The grain-density of a track in the upper 100 μ layer was found to be 0·45% higher, and in the bottom layer 0·26% lower than in the centre layer ; the statistical deviation for each layer was 1·0% (defined as $\sigma=100/\sqrt{N}$, the percentage standard deviation, where N is the number of grains counted). With this counting convention, there was therefore no significant variation of grain density with depth.

Using this result, a search was made for local variations in the scanned areas of plates *A*, *B*, *C* and *D*. Plate *A* had a grain-density 1·58% lower, *C*, 0·84% lower, and *D* 2·81% higher than plate *B* which was taken as standard. (For each measurement σ was 1·0%.) The difference of 4·39% between plates *A* and *D* is apparently significant. It is presumably either due to a variation of sensitivity of the poured emulsion, to non-uniform development, or to both factors. The variation between plates poured and developed at different times is likely to be even greater.

§3. ENERGY DEPENDENCE OF THE GRAIN DENSITY

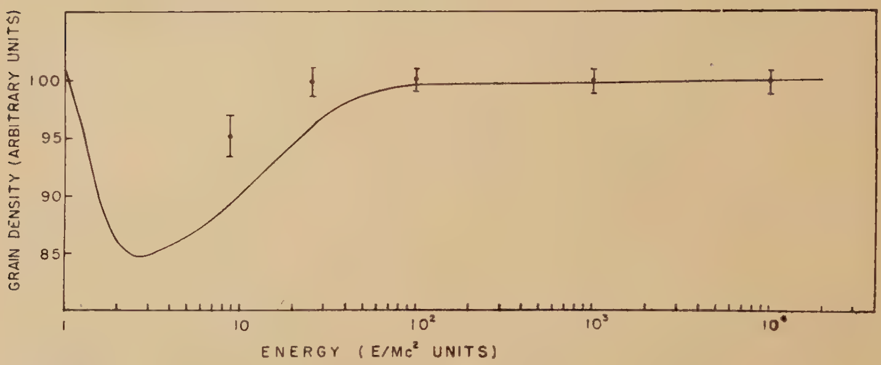
The grain-densities of tracks of electrons for the intervals of E/mc^2 of (1) 6–12 ; (2) 20–32 ; (3) 50–150 ; (4) 500–1 500 and (5) 5 000–15 000 were determined with the help of the preceding calibration. (σ was 1·8%, 1·3%, 1%, 1% and 1% respectively.) The results are plotted in fig. 2. The grain density is constant to within 0·25% for the energy ranges (2) to (5). The decrease of 4·8% for group (1) is apparently statistically significant. It may be noted, that, if the calibration were not applied, points (2) to (5) would still be constant to within 1·0% and the decrease for group (1) would be 5·5%. The results are in substantial agreement with the work of Pickup and Voyvodic (1950), which extends over a small range of values of E/mc^2.

The first detailed theoretical calculation of the ionization losses in condensed media was made by Fermi (1940). He showed how polarization effects modified the Bethe-Bloch theory, making the simplifying assumption that the dielectric properties of the medium were characterized by only a single dispersion frequency. Wick (1943) and Halpern and Hall (1948) put forward a multi-frequency theory. Bohr (1948) and Schönberg (1950) have also studied this problem and, in particular, have considered in some detail the effect of the Cerenkov radiation. All these theoretical treatments are entirely classical, but Allen (1951) has investigated the problem from a quantum standpoint.

Messel and Ritson (1950) pointed out that in the photographic plate, it was not the total energy absorbed per unit volume that should be considered, but rather only that part which was effective in latent image formation (Webb 1948). They therefore suggest that the contribution to the absorption of energy due to δ-rays of range greater than the mean grain-size, should be neglected. This is done by applying a ‘ cut-off ’ to the maximum energy equal to 5 or 10 kV.

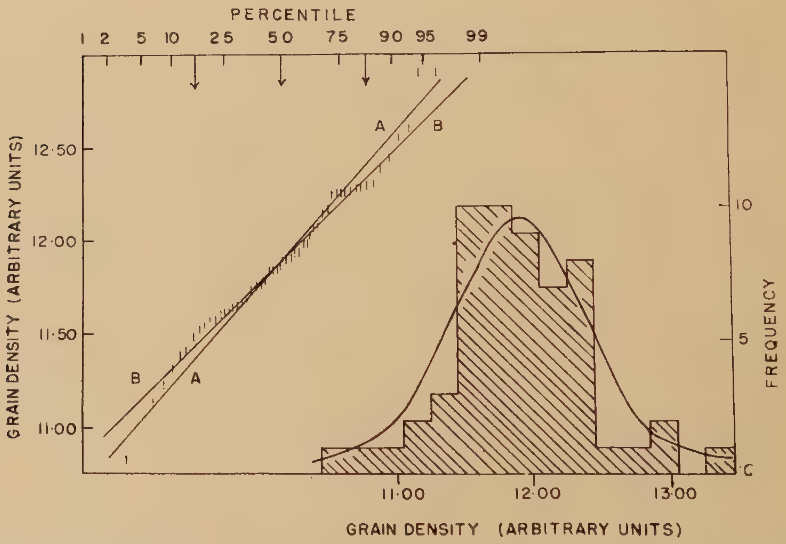
There is a lack of experimental data concerning the oscillator strengths and excitation energies for the electron transitions of silver bromide, particularly in the ultraviolet region. Theoretical calculations on the basis of the Hartree method would lead to only very approximate eigenfunctions because of the occurrence of the phenomenon of ‘ over stability ’. Consequently, no computations for the ionization loss in AgBr were made. In fig. 2, the curve is a calculation for iron, using a cut-off energy of 5 kV. The corresponding curve for AgBr can be expected to differ from this by a few per cent. However, the theory appears to be capable of accounting for the main features observed in this experiment.

Fig. 2



The observed variation of grain density with E/mc^2. The calculated curve is for iron, calculated from the theory of Halpern and Hall.

Fig. 3



Probability curves designed to investigate latent image fading. The experimental points are indicated with maximum reading errors. The grain density is measured per 50·6 microns of track length.

It would be of interest to investigate the grain density at $E/mc^2 \sim 2$ or 3, but even with an exposure to a homogeneous electron beam at this energy, the experiment is very difficult owing to the large scattering in the emulsion. A mono-energetic meson beam in this region is not yet available.

§4. FADING OF THE LATENT IMAGE

The tracks were examined in order to determine if there were appreciable fading effects in the G5 emulsion, due to any extreme conditions of temperature and humidity during the course of a high-altitude exposure. The histogram of fig. 3 shows the distribution in grain density normalized with the aid of the previous calibration, of 57 tracks of electrons whose energies were in the ranges (2) to (5). The length of each track, starting from the origin of the electron pair, was determined by the convention that it has $N=500$ grains. Since, for a normal distribution, 68% of the tracks should lie within $N+\sqrt{N}$ and $N-\sqrt{N}$, a broader distribution would imply that fading was appreciable. To test this point, the experimental data were plotted on probability paper (in which a normal distribution becomes a straight line), and the result is shown in fig. 3. The arrows indicate the 50th (the mean) and the 16th and 84th (the limits set by the statistical deviation) percentiles. *AA* is the calculated normal curve, while *BB* is a better fit for the points. Hodgson (1952) has pointed out that because of the finite grain size, the statistical fluctuations are somewhat reduced, an effect analogous to the counting losses of a Geiger-Müller counter due to finite resolving time. It may therefore be concluded, since there is no significant difference in slope between the curves *AA* and *BB*, that any effects due to fading were small and without significant influence on the results of this experiment.

ACKNOWLEDGMENTS

It is a pleasure to thank Professor C. F. Powell for the opportunity to study in this laboratory, and for his kind interest in this experiment. The author wishes to acknowledge helpful discussions with J. R. Allen, U. Camerini, J. E. Hooper, D. T. King and others at Bristol, and W. W. Happ at Vancouver. He is indebted to Mrs. D. M. Ford, Miss M. Jones and Miss J. Witchell for scanning the plates. This work was carried out at Bristol while the author was on a National Research Council of Canada post-doctorate fellowship.

REFERENCES

ALLEN, J. R., 1951, Private Communication.
BOHR, A., 1948, *K.D.V.S.* 24, No. 19.
FERMI, E., 1940, *Phys. Rev.*, **57**, 485.
GOTTSTEIN, K. L. F., MENON, M. G. K., MULVEY, J. H., O'CEALLAIGH, C. and ROCHAT, O., 1951, *Phil. Mag.*, **42**, 708.
HALPERN, O., and HALL, H., 1948, *Phys. Rev.*, **73**, 477.
HODGSON, P. E., 1952, *Brit. J. Appl. Phys.*, **3**, 11.
HOOPER, J. E., KING, D. T., and MORRISH, A. H., 1951, *Can. J. Phys.*, **29**, (in press).
MESSEL, H. and RITSON, D. M., 1950, *Phil. Mag.*, **41**, 1129.
SCHÖNBERG, M., 1950, *Bulletin du Centre de Physique Nucleaire de L'Universite Libre de Bruxelles, No.* 20, August.
WEBB, J. H., 1948, *Phys. Rev.*, **74**, 511.
WICK, C. G., 1943, *Nuovo Cim.*, **9**, 3, 303.

XLVIII. *The Origin of Glauert's Superposition Fringes*

By S. C. Hunter and F. R. N. Nabarro
Department of Metallurgy, University of Birmingham *

[Received February 11, 1952]

Summary

Glauert has recently published photographs showing the distortion of multiple-beam interference fringes produced by inserting a thin sheet of glass or mica between two interferometer mirrors. It is shown that multiple reflections between the surfaces of the inserted sheet and the mirrors necessarily produce the types of distortion observed, and that the published photographs can be analysed quantitatively.

§1. Introduction

This paper is concerned with the interpretation of some remarkable interference fringes recently published by Glauert (1951). 'The Fizeau fringes given by a pair of flats, placed at a small angle to each other, are straight, lying parallel to the edge of the wedge. If the flats are silvered and the monochromatic illumination is correctly collimated, the fringes are formed by multiple reflections and are very sharp . . . It is often proposed to place a (biological) specimen under examination in a suitable fluid between silvered optical flats, in the hope that the resultant fringes will give some indications as to structure and form. The present results show that the greatest care must be taken in the interpretation of such interferograms, and they illustrate the ease of production of artefacts.' The simplest results to interpret are those shown in figs. 1 and 2 of Glauert's paper which are reproduced in Plate XXII (*a*) and (*b*).† In Plate XXII (*a*), a glass cover slip had been inserted between the flats, and the resulting fringes were modulated into zigzags, the amplitude of the zigzag itself varying along each fringe. The positions of minimum amplitude did not correspond on neighbouring fringes, but were approximately repeated every third fringe. In Plate XXII (*b*), a sheet of mica had been inserted between the flats. There was a region of violent disturbance at the left of the picture, which we do not attempt to interpret. Elsewhere the fringes were doubled, and the two components could be separated by using polarized light. In one component the amplitude of the zigzags was uniformly small, while in the other it was uniformly large. Except on two fringes near the middle of the field, the zigzags of large amplitude were very unsymmetrical, actually overhanging in opposite directions in the fringes on opposite sides of these two.

---

* Communicated by the Authors.

† For plate see end of issue.

As Glauert suggests, the observed effects can all be explained in terms of multiple reflections between the mirrors and the interposed sheet, and between the two faces of this sheet. The optical path for a simple passage from the upper mirror to the lower is, of course, independent of the position of the interposed sheet. In practice a ray cannot make this simple passage, but parts of it suffer single or multiple reflections from the surfaces of the sheet. The amplitude of the resultant ray depends on the phases of those disturbed components relative to the component which has made the simple passage, and therefore on the position of the sheet between the mirrors. If the sheet is tilted about an axis lying perpendicular to the beam of light and to the intersection of the mirrors, the effective optical thickness is not constant along one of the original fringes, and the fringes observed can no longer be straight lines parallel to the intersection of the mirrors.

For a detailed interpretation, we have to make the reasonable assumptions that the glass slip was very nearly plane, but that its thickness varied gradually by a few fringes in crossing the field, while the slip of mica was of very uniform thickness, but appreciably twisted out of a plane.

In § 2 we discuss the multiple reflection of a single ray between two mirrors with a sheet of material of thickness T and refractive index μ lying between them. In § 3 we discuss the fringe pattern formed by a plane parallel sheet lying obliquely between the mirrors, and in § 4 we consider the additional effects which arise if the faces of the sheet are slightly inclined to one another to form a wedge.

Our interpretation of these fringes suggests that the corresponding artefacts will be much less noticeable in the case of a biological specimen in a fluid of refractive index close to its own than in the present simple case. Firstly, the lines of maximum rate of change of refractive index will in general make finite angles with the primary beam of light, and the reflected components will therefore not interfere with the direct beam. Secondly, the amplitudes of the internal reflections will be small if violent changes of refractive index are avoided.

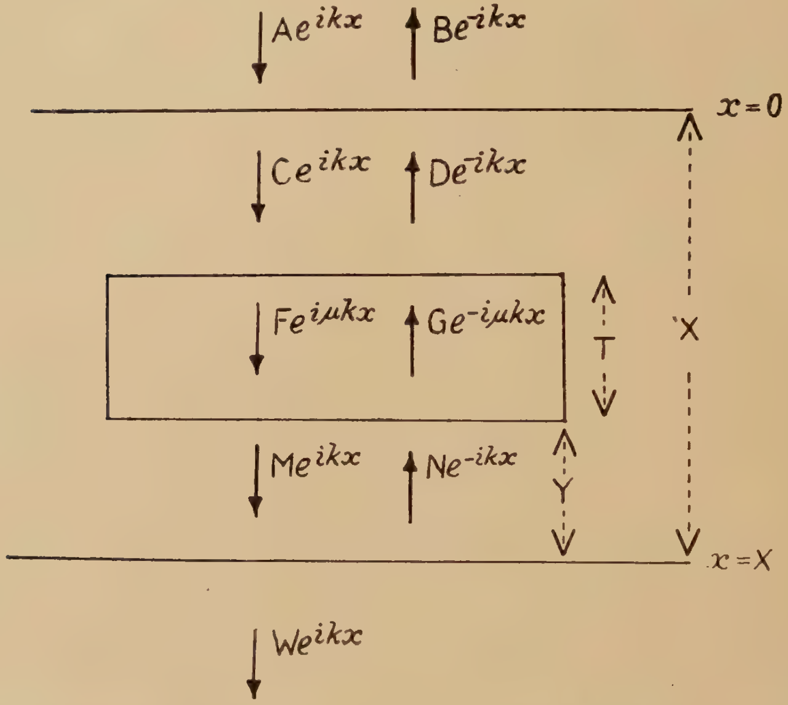
§2. MULTIPLE REFLECTIONS OF A SINGLE RAY

We consider an incident ray with electric component $A \exp(ikx)$ striking the upper silvered surface ($x=0$), suffering multiple reflections, and finally emerging as a reflected ray $B \exp(-ikx)$ and a transmitted ray $W \exp(ikx)$. Between the upper silvered surface and the upper surface of the glass plate, the reflected rays combine to form resultants $C \exp(ikx)$ propagating downwards, and $D \exp(-ikx)$ propagating upwards, and similarly in the other two regions (fig. 1). The refractive index of the glass is μ, and the distances X, Y and T are indicated in the figure.

Let t and r be the coefficients of transmission and reflection at a glass-air interface, defined so that the transmitted flux of energy is $|t|^2$ times the incident flux, while the phases of the electric vector in the transmitted and reflected waves are given correctly by t and r. Then for a surface at $x=0$

$$\left.\begin{array}{l} t=2\mu^{1/2}/(\mu+1), \\ r=\pm(\mu-1)/(\mu+1), \\ \text{the sign depending on the direction of the incident ray,} \\ \text{and} \qquad r^2+t^2=1. \end{array}\right\} \quad (1)$$

Fig. 1



Multiple reflections of a single ray. The incident ray A gives rise to a reflected ray B and a transmitted ray W.

The corresponding quantities for the silvered mirrors, t' and r' are in general complex, and $|t'|^2+|r'|^2 \leqslant 1$. We can, without loss of generality, take the mirrors to be thin silvered sheets with no glass backing.

The continuity of **E** at each interface leads to the equations

$$\left.\begin{array}{r} B=r'A+t'D, \\ C=t'A+r'D, \\ De^{-ik(X-T-Y)}=-rCe^{ik(X-T-Y)}+\mu^{1/2}tGe^{-i\mu k(X-T-Y)}, \\ Fe^{i\mu k(X-T-Y)}=\mu^{-1/2}tCe^{ik(X-T-Y)}+rGe^{-i\mu k(X-T-Y)}, \\ Ge^{-i\mu k(X-Y)}=rFe^{i\mu k(X-Y)}+\mu^{1/2}tNe^{-ik(X-Y)}, \\ Me^{ik(X-Y)}=\mu^{1/2}tFe^{i\mu k(X-Y)}-rNe^{-ik(X-Y)}, \\ Ne^{-ikX}=r'Me^{ikX}, \\ \text{and} \qquad W=t'M. \end{array}\right\} \quad (2)$$

Here r stands for $+(\mu-1)/(\mu+1)$.

Solving these equations for the ratio W/A, we find

$$|W/A|=t'^2t^2/\eta, \qquad \ldots\ldots\ldots \quad (3)$$

where η is a rather complicated expression. We know that as the thickness of silvering on the mirrors is increased, so that $|r'|\to 1$ and $|t'|\to 0$, the interference pattern becomes fainter, while the bright fringes become narrower. In the limit $|r'|\to 1$, the pattern disappears completely, but the positions of these fringes are given by the condition $\eta=0$, which can now occur for real values of X and Y. By adding constants to X and Y, we may take $r'=+1$ for both mirrors. The condition $\eta=0$ then becomes

$$\begin{aligned}&(1+re^{-2ikY})-re^{-2i\mu kT}(r+e^{-2ikY})\\&=e^{-2ik(X-T-Y)}[e^{-2i\mu kT}(r+e^{-2ikY})-r(1+re^{-2ikY})]. \qquad \ldots \quad (4)\end{aligned}$$

The ray considered contributes to a bright fringe if this condition is satisfied, and if X, Y and T are functions of position, the equation of a fringe is

$$\left.\begin{aligned}&X-T=\rho'(Y)+(2n+1)\pi/2k\\ \text{where}\qquad &\tan k\rho'=\frac{(1-r^2)\cot\mu kT+2r\sin 2kY}{(1+r^2)+2r\sin 2kY}\end{aligned}\right\}, \qquad \ldots\ldots \quad (5)$$

which may also be written

$$\left.\begin{aligned}&X+(\mu-1)T=\rho(Y)+n\pi/k\\ \text{where}\quad &\tan k\rho=-\frac{r^2\sin 2\mu kT+2r\sin\mu kT\cos(2kY+\mu kT)}{1-r^2\cos 2\mu kT+2r\sin\mu kT\sin(2kY+\mu kT)}\end{aligned}\right\} \quad \ldots \quad (6)$$

Here n is any integer.

§3. FRINGES FORMED BY INCLINED MIRRORS WITH A SHEET OF UNIFORM THICKNESS INTERPOSED

We now suppose that the various surfaces in fig. 1 are not exactly parallel, although so nearly parallel that multiply reflected beams will combine to give interference. In the plane of the lower mirror we choose axes u, parallel to the intersection of the mirrors, and v, perpendicular to this (fig. 2). Then X is a linear function of v

$$X=\theta v. \qquad \ldots\ldots\ldots\ldots \quad (7)$$

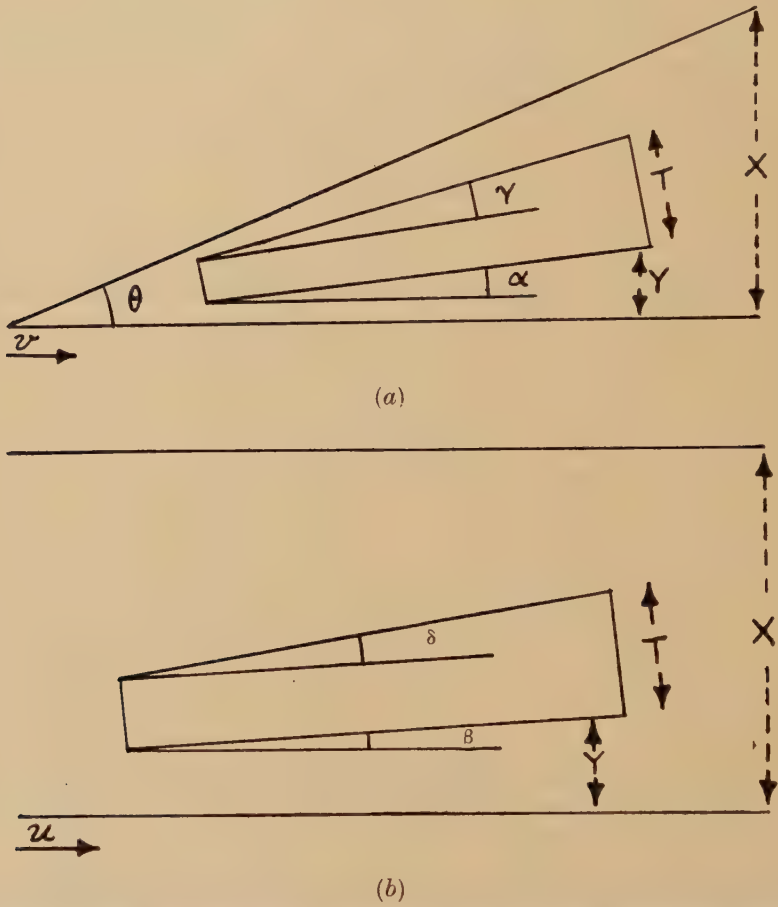
If we take $\mu=1$, $r=0$, the fringes are given by (6) as

$$\left.\begin{aligned}X&=n\pi/k,\\ \rho&=0.\end{aligned}\right\} \qquad \ldots\ldots\ldots\ldots \quad (8)$$

Whatever the value of r, the numerator in the expression for $\tan k\rho$ is bounded, while the denominator obviously cannot vanish if $1-r^2-2r>0$ i.e. $r<\sqrt{2}-1, \mu<\sqrt{2}+1$. (The denominator is, in fact, always positive if $r<1/\sqrt{2}$, $\mu<3+2\sqrt{2}$.) For any practical value of μ, $\tan k\rho$ remains bounded as μ increases from 0 to its actual value. If we number the fringes in the absence of a glass plate by the integer n, we may establish a 1 : 1 correspondence between the fringes and those in the presence of a glass plate by adding to (6) the condition

$$-\tfrac{1}{2}\pi < kp < \tfrac{1}{2}\pi.$$

Fig. 2



Sections showing the glass plate between the two mirrors. (*a*) section normal to the intersection of the mirrors, (*b*) section parallel to the intersection of the mirrors and to the rays of light.

Suppose now that the glass plate is not parallel to the intersection of the mirrors, so that Y is a function of u. It follows from (5) or (6) that the fringes are no longer straight lines, v=constant, but are modulated in the v direction. For a given thickness T, the amplitude can be determined from (5). If we put $\tan kY=Z$, there results a quadratic equation in Z, the roots of which coincide if

$$\tan kp' = \frac{1+r^2}{1-r^2}\cot \mu kT \pm \frac{2r}{1-r^2}\operatorname{cosec} \mu kT. \quad . \quad . \quad . \quad (9)$$

If we call the corresponding values ρ_1' and ρ_2', we have

$$\tan k(\rho_1'-\rho_2')=2\phi/(1-\phi^2),$$

where

$$\phi=2r\sin\mu kT/(1-r^2).$$

Since the extreme values of ρ' must coincide if $r=0$, the appropriate solution is

$$k|\rho_1'-\rho_2'|=2\tan^{-1}|2r\sin\mu kT/(1-r^2)|. \quad . \quad . \quad . \quad (10)$$

This gives the total amplitude of displacement of a fringe $|\rho_1'-\rho_2'|$ on the scale for which the distance between neighbouring fringes is π/k. (The same formula (10) may be obtained from (6): there is then no ambiguity in the choice of solution.)

In the case of the glass cover slip, the thickness varies appreciably in the field, and the amplitude varies from zero to its maximum value of $(2/\pi)\tan^{-1}|2r/(1-r^2)|$ times the spacing of fringes. By measuring the maximum amplitude and the spacing, we find $2r/(1-r^2)=\tan 26{\cdot}2°$, giving $\mu\simeq 1{\cdot}61$, a reasonable value. In the case of mica, the thickness is apparently constant to a small fraction of a fringe. This thickness is such that for one polarization $\sin\mu kT$ is practically zero. If we assume that for the other polarization $\sin\mu kT$ is unity (so that the mica formed a quarter-wave plate), we find $\mu=1{\cdot}57$, while the principal refractive indices of mica lie between 1·56 and 1·60.

We now consider in more detail the effect of inserting a slightly tilted plate of uniform thickness. In addition to eqn. (7) for X, we have

$$Y=Y_0+\alpha u+\beta v, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

where α and β are small angles comparable with θ. The equation of the fringes in the (u, v) plane is given by (6) as

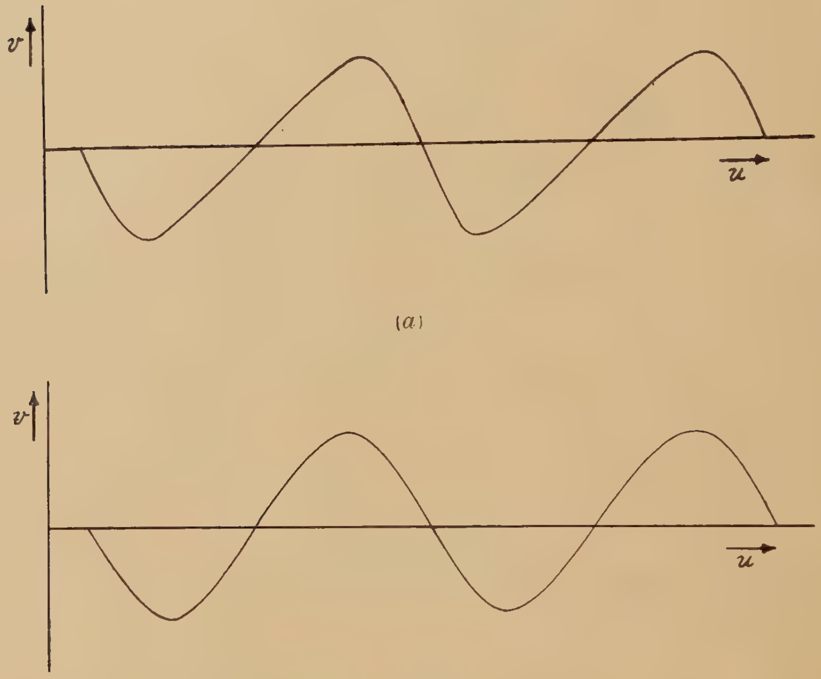
$$\theta v+(\mu-1)T=\rho(Y_0+\alpha u+\beta v)+n\pi/k \quad -\tfrac{1}{2}\pi<k\rho<\tfrac{1}{2}\pi. \quad . \quad . \quad (12)$$

It follows from (6) and (12) that a given fringe undergoes one complete cycle of modulation when $k\alpha u$ increases by π, while we pass from one fringe to the next when $k\theta v$ increases by π. The observed ratio of the distance between fringes to the wavelength of the modulation gives α/θ directly: for the glass cover slip we find $\alpha/\theta=1{\cdot}73$. For the mica, the passage from fringes of short wavelength at the top of the figure to those of long wavelength at the bottom indicates that the sheet was warped, dipping (or rising) steeply to the right near the top, and less steeply near the bottom.

The form of the modulation, given by (6), is noticeably unsymmetrical. It is sketched in fig. 3 (*a*) for the case $\sin\mu kT=1$, $\mu=1{\cdot}61$.

The effect of the term βv is to shear the whole pattern through an angle $-\beta/\alpha$, the lines $v=$constant remaining parallel to themselves (' simple shear '). As is shown in fig. 4, this shear may turn the original modulations into overhanging teeth. This is actually observed in Plate XXII (*b*). The fringes at the top of the figure overhang to the left, those in the middle are symmetrical, and those at the bottom lean to the right. We deduce that the mica was bent about an axis roughly parallel to the fringes. The fringes observed with glass do not overhang, and are, in fact, more nearly symmetrical than those of fig. 3 (*a*). We deduce that they have been sheared through an angle of about $\tan^{-1}(0{\cdot}4)$ when their form should be as in fig. 3 (*b*). It follows that $\beta/\theta \simeq 0{\cdot}7$.

Fig. 3



(*a*)

(*b*)

Calculated shape of the fringes of maximum amplitude, with refractive index $\mu=1{\cdot}61$. (*a*) undistorted, (*b*) sheared through an angle of $\tan^{-1} 0{\cdot}4$.

§ 4. Effect of a Gradual Variation in the Thickness of the Sheet

We shall now show that three other effects observed in the fringes formed with a glass cover slip can be explained on the assumption that the thickness of the glass T was not constant, but varied linearly in the field of the picture according to the law

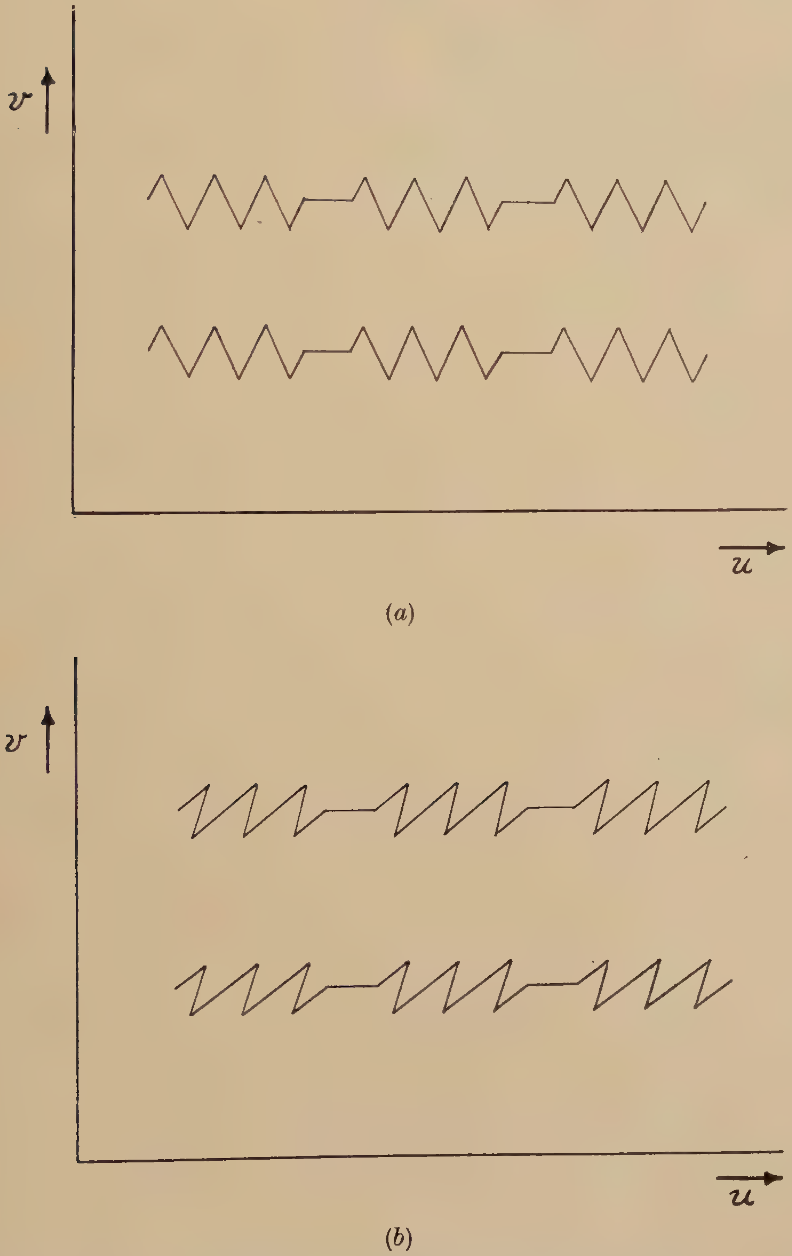
$$T=T_0+\gamma u+\delta v. \qquad (13)$$

These effects are (i) that the mean spacing of the fringes where the glass is interposed is not equal to the mean spacing of the undistorted fringes, (ii) that the amplitude of modulation is not constant along any one fringe, and (iii) that points of minimum amplitude on neighbouring fringes do not correspond, while those on every third fringe usually correspond.

We again assume that T varies so slowly that the glass in any neighbourhood may be considered as of uniform thickness. The equation of the fringes is then

$$\theta v+(\mu-1)(T_0+\gamma u+\delta v)=\rho+n\pi/k \quad . \quad . \quad . \quad . \quad (14)$$

Fig. 4



Effect of the shear produced by the angle of tilt α (schematic). (*a*) fringes before shearing, (*b*) fringes after shearing.

and the condition $-\frac{1}{2}\pi < k\rho < \frac{1}{2}\pi$ ensures that the term in ρ does not affect the mean direction or spacing of the fringes.

The mean direction of the fringes is such that $[\theta+(\mu-1)\delta]v+(\mu-1)\gamma u$ is constant, while the direction of the undistorted fringes is v=constant. Thus the mean direction of the fringes turns on meeting the glass through an angle $\tan^{-1}\{(\mu-1)\gamma/[\theta+(\mu-1)\delta]\}$. This cannot be observed in the picture. The mean spacing of the fringes in the v direction is $\pi/k[\theta+(\mu-1)\delta]$. In the absence of the glass the mean spacing is $\pi/k\theta$. The ratio of these spacings is $1+(\mu-1)\delta/\theta$: the observed ratio is 8/7, and we have determined $\mu=1{\cdot}61$, so that $\delta/\theta=0{\cdot}2$.

The variation of amplitude along a fringe is given by (10). The modulation vanishes whenever $\mu kT=n\pi$. If the points (u_1, v_1) and (u_2, v_2) lie on the same fringe, we have from (14) that

$$\theta(v_1-v_2)+(\mu-1)[\gamma(u_1-u_2)+\delta(v_1-v_2)]=0,$$

while if they are neighbouring points of zero modulation we have

$$\mu k[\gamma(u_1-u_2)+\delta(v_1-v_2)]=\pi.$$

Eliminating v, we find

$$(u_1-u_2)=[1+(\mu-1)\delta/\theta]\pi/\mu k\gamma.$$

The ratio of this to the mean spacing of fringes in the v direction is thus

$$[1+(\mu-1)\delta/\theta]^2\theta/\mu\gamma.$$

The observed ratio is about 11. Since $[1+(\mu-1)\delta/\theta]=8/7$, we have $\gamma/\theta=0{\cdot}076$. The angle through which the fringes should be bent on meeting the glass is thus $\tan^{-1} 0{\cdot}04 < 2\frac{1}{2}°$, and it is reasonable that this is not observed.

Finally we check this portion of the analysis by means of the observation that the positions of zero amplitude on roughly every third fringe correspond. For we go from one zero to another by moving in the v direction a distance $\pi/\mu k\delta$, which is $(\theta/u\delta)[1+(\mu-1)\delta/\theta]$ fringes. Using the values of δ/θ and μ already determined, this is equal to 3·03 fringes, agreeing satisfactorily.

ACKNOWLEDGMENTS

One of us (S. C. H.) acknowledges the financial support of the Department of Scientific and Industrial Research. We are grateful to Miss Glauert for helpful suggestions.

REFERENCE

GLAUERT, AUDREY M., 1951, *Nature, Lond.*, **168**, 861.

XLIX. *An Unusual Birefringence Interference Pattern*

By S. Tolansky and F. S. A. Sultan
Royal Holloway College, Egham*

[Received February 11, 1952]

In a recent note to *Nature*, Glauert (1951, *Nature, Lond.*, **168,** 861–2) has described, but did not explain, some unusual localized interference fringes obtained by placing a thin sheet of mica between two suitably silvered optical flats. On viewing with monochromatic radiation a highly complex system of localized multiple beam fringes form, with modulations arising through the wrinkling of the mica which, in fact, leads to a triple interference film, two air gaps and one mica gap. A striking feature of the fringes described by Glauert is the anomaly arising from the birefringence of the mica. The fringes are double, the two components being mutually perpendicularly plane polarized, but they do not follow identical modulation patterns, being variously straight or zig-zag according to location, even at times when very close to each other. The modulation effect is clearly determined both by the geometry and the refractive index.

We wish to describe here another unusual interference phenomenon which is closely allied to and which throws some light on the behaviour of the fringes described by Glauert. A drop of molten para-toluidine was placed between two silvered optical flats and allowed to crystallize *in situ.* The toluidine contacted the surfaces and therefore there were no effective air films. On examination with mono-chromatic light (green mercury) the system showed a distribution of multiple beam Fizeau fringes which were double owing to the crystal birefringence. One small region which is illustrated in fig. 1 (*a*) (Plate XXIII†) exhibited an anomaly which we have not seen elsewhere in any of the many hundreds of interferograms recorded in this laboratory. The fringe pattern consisted of a network of two systems mutually plane polarized perpendicularly, these two systems crossing each other almost at right angles. Fig. 1 (*a*) shows the fringe pattern whilst 1 (*b*) and 1 (*c*) show the two patterns obtained respectively by the introduction of polaroids, set, in the two cases, at right angles to each other. This system of two independent sets of fringes can only be interpreted one way. The crystal is acting as a simple wedge to each direction of polarization, but the two ‘ wedges ’ are almost at right angles. One is confronted with a position in which there is a geometrical wedge in one direction and an optical wedge in another direction. The geometrical wedge forms because the surfaces of the flat are not parallel.

---

* Communicated by the Authors.

† For plate see end of issue.

The crystal axes must be so oriented in this particular form of crystallite such that with respect to the two different polarizations the effective wedges are roughly at right angles to each other.

This curious fringe pattern appears to be related in form to those described by Glauert, but much simplified through the absence of the air films, and the wrinkling.

L. *On the Significance of the Limit Load Theorems for an Elastic-plastic Body**

By E. H. Lee
Graduate Division of Applied Mathematics, Brown University, U.S.A.†

[Received January 15, 1952]

Abstract

The limit load theorems of Drucker, Greenberg, and Prager (1951) for an elastic-plastic material flowing under constant surface tractions are contrasted with the theory of the yield of a plastic-rigid body (Hill 1951). Plastic-rigid theory is interpreted, as it must be for application to actual materials, as a model providing an approximation which can reasonably be expected to be satisfactory when plastic strains large in comparison with elastic strains have occurred. It is shown by the discussion of particular examples that the limit load for an elastic-plastic body is in general reached or closely approached while strains throughout are of the order of elastic strains, and at this stage the limit load theorems can be applied. Examples are cited to show that if plastic-rigid analysis is adopted, and strains large compared with elastic strains must be attained, the influence of boundary motion may invalidate application of the yield limit theory in circumstances when the limit load theorems for an elastic-plastic material provide significant results.

It is shown that the uniqueness and variational theorems applicable to a plastic-rigid material at yield, apply to an elastic-plastic material at the limit load, thus permitting stress distributions in an elastic-plastic body to be determined on the basis of plastic-rigid theory. Examples are cited.

§1. Introduction

We are concerned with bodies of elastic ideally-plastic material, having a stress-strain relationship of the type shown in fig. 1. Quasi-static motion is considered for which inertia forces can be neglected. When such bodies are subjected to external tractions, as a rule they are initially stressed elastically throughout. With increasing tractions plastic flow will commence, but there will in general be a stage of contained plastic flow, in which the regions flowing plastically are constrained by surrounding elastic material so that strains throughout are of the order of elastic strains. With increasing loads, the plastic regions will in general spread until such a constraint ceases to apply, and continued flow takes place with little or no increase in load. This situation is termed uncontained flow, and the constant load at which continued flow can occur is termed the limit load. Recently Drucker, Greenberg and Prager (1951) have developed theorems prescribing whether or not such continued flow will take place under specific boundary conditions (see also Prager and Hodge 1951, chap. 7). When particular situations are considered, these theorems can be used to determine upper and lower bounds to the limit loads. The power of these theorems lies in the circumstance that they provide information about the uncontained flow, without the necessity of analysing the previous contained flow which is in general beyond the scope of presently available methods of analysis.

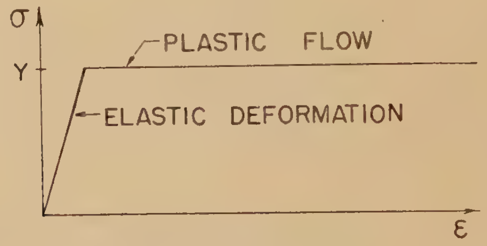
---

* The results presented in this paper were obtained in the course of research sponsored by the Office of Naval Research.

† Communicated by the Author.

Recently Hill (1951) has discussed the state of stress at the yield point of a rigid-plastic body, that is, from the standpoint of application, a body in which elastic strains are neglected so that the elastic line in fig. 1 is considered to lie on the σ axis, deviation from it occurring only at the yield point stress Y. Such a hypothesis can only be expected to provide a satisfactory approximation for an actual elastic-plastic material if the plastic strains occurring are large compared with elastic strains, so that the latter can be neglected compared with the former. For a rigid-plastic material, no deformation occurs until uncontained flow takes place, and Hill presents theorems concerning this initial yield which are in many respects similar in form to the earlier theorems of Drucker, Greenberg and Prager for an elastic-plastic body. Hill's theorems are obtained by a generalization and reinterpretation of his work, Hill (1948, 1950), on the principle of maximum work and its complement, and he states that the theorems of Drucker, Greenberg and Prager are special cases of this generalized principle. To show that this is not the case is one of the objects of the present paper, and it is demonstrated in the next section that in certain cases it is essential to consider the deeper elastic-plastic analysis in order to be able to make useful application of such theorems in practice.

Fig. 1



Stress-strain relation.

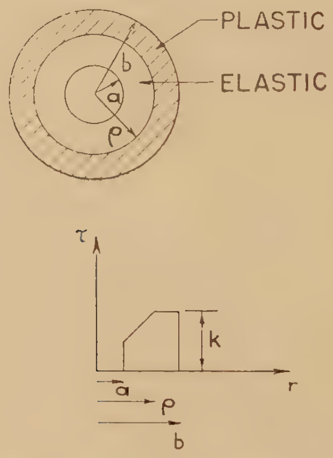
The reason for this difficulty is that in the application of both groups of theorems, the initial unstressed surface of the body is used to specify boundary conditions. For a plastic-rigid body at yield, previous strains being of the order of elastic strains are logically neglected, and the undeformed surface is therefore used. For the elastic-plastic body the theorems are expressed in terms of the current surface geometry when uncontained flow is impending, but since in general the contained flow has not been analysed, the deformation of the boundaries is not known, and so the initial undeformed boundaries must be used as an approximation to the current boundary geometry. For many problems involving strains of the order of elastic strains the motion of the boundaries has a negligible influence, in fact the usual linear theory of elasticity is based on this hypothesis. However, as the slenderness of the geometrical form increases, boundary motion has an increasingly important influence, and for example in the case of elasticity we reach the group of elastic instability problems for which boundary motion dominates the solution. The influence of boundary motion clearly becomes more important as the strains increase, so that application of a rigid-plastic analysis, which can only be expected to give useful results for an actual body when strains occur which are large compared with elastic strains, may imply a marked influence of boundary motion. An analysis neglecting this will then be unsatisfactory. In many cases the elastic-plastic analysis leads to attainment or close approach to the limit load while the strains are still of elastic order, say twice the elastic limit strains, so that the limit load theorems can be used to give significant results before boundary motion has an appreciable influence. Thus the limit load for an elastic-plastic body may be significant, whereas the practical application of the corresponding plastic-rigid analysis which requires strains large compared with elastic strains may be vitiated by the influence of boundary motion. Examples are given in the next section to illustrate this situation quantitatively.

Before turning to the discussion of specific examples let us consider the first theorem of Drucker, Greenberg and Prager (1951) which contains the basic difference between their analysis and that of Hill. It states that if changes in geometry which occur during uncontained flow under constant surface tractions are neglected, then the stresses remain constant, and so only increments in the plastic components of strain occur. Thus the elastic strain increment components are zero, and it is for this reason that the resulting theorems for an elastic-plastic body bear a close resemblance to corresponding theorems for a rigid-plastic body in which from the standpoint of practical application the elastic components of strain are simply neglected in comparison with the plastic components.

§2. BEHAVIOUR IN UNCONTAINED FLOW

Without a detailed knowledge of each particular case the term uncontained plastic flow, that is flow not constrained by surrounding elastic material, is difficult to define precisely. From the study of particular examples it is found that sometimes the spread of the plastic region reaches a stage at which continued plastic flow occurs under constant surface tractions, if geometry changes are neglected, while the strains throughout are of the order of elastic strains. In other cases, although there is no direct constraint due to surrounding elastic material, for continued flow the surface tractions must increase towards a limiting value, and in general almost attain this value while strains throughout are still of elastic order of magnitude. Examples of such behaviour have appeared in the literature (e.g. Hill, Lee and Tupper 1947), and such situations have recently been emphasized by Hill (1951). The conditions under which this increase occurs may be thought of as a type of constraint due to the elastic properties of the material, for in the notation of the theory covering a particular example given in the Appendix, it represents a lag in the orientation of the stress vectors behind the strain increment vectors which prevents exact 'free' flow, that is flow under constant stress. Such a lag is due to the elastic components of strain. Thus generalizing our concept of elastic constraint in this way, limit flow occurs under constant or almost constant surface tractions, and to this extent the limit load theorems are exact or nearly so. From the study of particular examples it is clear that in general for reasonably steady boundary conditions this lack of preciseness in the attainment of the limit load concerns differences of very small magnitude which are of no importance in the practical application of the theorems.

Fig. 2



Stress distribution in twisted tube.

Since only a few elastic-plastic problems are amenable to complete analysis, we can use only simple problems for illustration. However, a quantitative study of these can be used to indicate the type of behaviour to be expected in more complex cases. As an example of flow at constant limit load, consider the torsion of a thick-walled tube of internal radius a, external radius b. When a torque below the elastic limit is applied, the shear stress is proportional to the radius and the torque increases linearly with the twist. When the shear stress at the outside reaches the limit value for plastic flow, k, plastic flow commences there, and with increasing torque the plastic region spreads towards the centre, the elastic-plastic boundary being a cylindrical surface concentric with the tube surfaces. Figure 2 shows the distribution of the shearing stress when the plastic elastic boundary is of radius ρ. With increasing torque the plastic region spreads, and the stress at each point in the plastic region remains constant. The plastic-elastic boundary reaches the interior surface at a twist b/a times the twist at the elastic limit, and thereafter the tube is fully plastic. While part of the tube is still elastic, increasing torque is required to produce increasing twist, but when the tube becomes fully plastic no further increase in torque is needed. We then have limit load conditions and stress rates are zero throughout the tube. For this particular case the deformation in the form of twist does not alter the shape of the tube, so that we have a special case in which geometry changes do not influence the limit load. The limit load theorems of Drucker, Greenberg and Prager will therefore apply exactly as soon as the tube becomes fully plastic, which for moderate ratios of b/a will be when the maximum strain is still of elastic order.

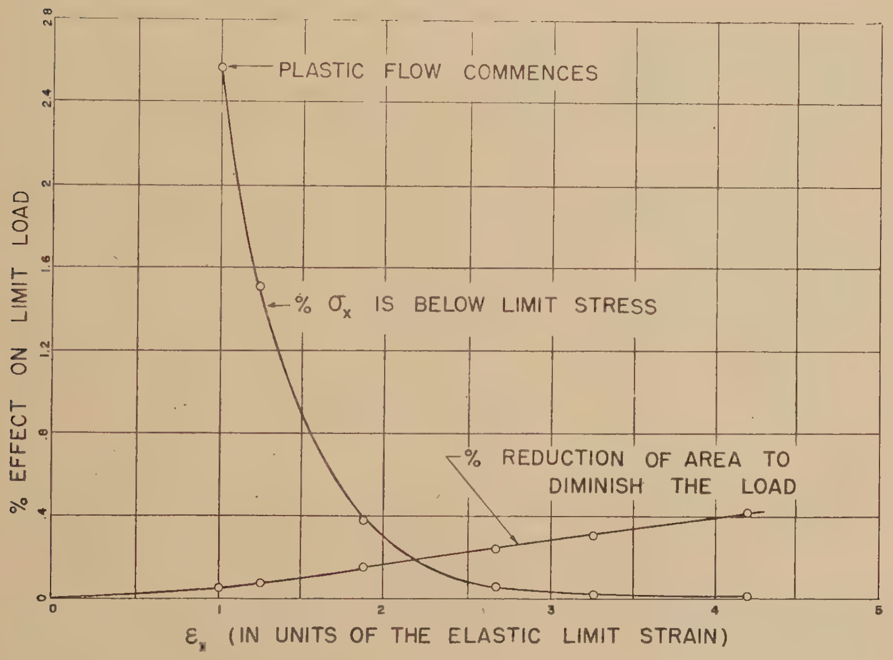
As an example for which, as far as elastic constraint is concerned, the limit load theorems apply exactly for strains of elastic order of magnitude, but which involves a boundary geometry influence, let us consider the expansion of a spherical shell under internal pressure. This problem has been fully discussed by Hill (1950 b, chap. V). Lamé's well known solution applies for internal pressure below the elastic limit. Plastic flow commences on the inside and at any stage the plastic-elastic boundary is a spherical surface concentric with the inner and outer surfaces. With increasing internal pressure a fully plastic state is reached when the plastic-elastic boundary reaches the outside surface, and thereafter continued flow occurs under constant internal pressure if boundary motions are neglected. Thus the conditions of the limit load theorems apply as soon as the spherical shell becomes fully plastic.

In fact, because of symmetry, the stress is given by the two components, radial and circumferential, which are statically determined by the equation of equilibrium and the yield condition when the boundaries are prescribed. Taking into account the influence of boundary changes, the internal and external radii increase due to the internal pressure, the thickness of the shell is reduced and continued fully plastic flow is deduced under decreasing internal pressure. For moderate ratios of internal to external radius the above limit load conditions pertain while the strains throughout are of the order of elastic strains. In practice of course, once the fully plastic condition arises with a non-hardening material, flow under falling pressure would be unstable and local thinning would be expected to take place, but the condition immediately following the development of full plasticity will be correctly analysed by the elastic-plastic limit load theorems. Adopting a strictly plastic-rigid analysis, one cannot be assured that deductions have any validity for an actual material until strains large compared with elastic strains occur, and at such a stage it is to be expected that local instability will have appreciably modified the solution. Thus whereas the elastic plastic approach provides definite information, nothing can be deduced with certainty from the plastic-rigid theory. It is this opportunity to apply the theorems of Drucker, Greenberg and Prager much earlier in the development of plastic flow, than the plastic-rigid theorems of Hill, which permits the much wider use of the former, in cases where the latter on their own cannot be expected to give applicable results. However, it is clear from the elastic-plastic analysis, that Hill's theorems, which are formally identical with those for the elastic-plastic body at the limit load, do in fact apply much earlier in the development of plastic flow than can reasonably be expected from a plastic-rigid analysis. This arises since elastic strain increments become negligible compared with plastic strain increments long before the total elastic strain can be neglected compared with the total plastic strain. Such an application of plastic-rigid theory has been made by Drucker, Greenberg, Lee and Prager (1951) to obtain information concerning boundary value problems for an elastic-plastic body. It is shown there that by substituting stress and strain distributions based on plastic-rigid analysis, the limit load theorems for an elastic-plastic body determine the exact limit load for this case. Thus plastic-rigid theory is shown to have a specific significance in elastic-plastic analysis, quite apart from the approximation reasonably to be expected for an elastic-plastic solution involving large strains.

Let us now consider an example in which the limit load is approached but not reached exactly, even neglecting the effects of boundary motion. We consider a tensile test in plane strain in which uniform tension σ_x produces uniform stress in a rectangular block of material, motion is prevented in the z direction so that ϵ_z is zero, and the faces normal to the y direction are free so that σ_y is zero. The resulting stress-strain relation has been given by Hill (1950 b, chap. IV) but in order to facilitate the assessment of the approach to uncontained flow, a solution is given in the Appendix based on the plane stress and strain diagram introduced by Reuss (1930) and discussed and utilized by Hill, Lee and Tupper (1947). This solution shows that with increasing tension σ_x, the block is initially entirely elastic before the elastic limit is reached at which the entire body becomes plastic. However, because of the difference in Poisson's ratio for elastic and plastic strains, the stress components at this stage are not compatible with free plastic flow in plane strain, and as this state is approached σ_x rises, and with increasing strain it rapidly approaches an asymptotic value, the limit stress. This is shown in fig. 3 where the percentage disparity of σ_x below the limit stress is plotted against the strain in units of the elastic limit strain. It is seen that σ_x is within 0·1% of the limit stress when the strain is only $2\frac{1}{2}$ times the elastic limit strain. Taking ϵ_x to represent the time, the rate of increase of surface traction becomes insignificant at this stage, and the elastic-plastic limit load theorems apply to a high degree of accuracy according to the development leading to Theorem I of Drucker, Greenberg and Prager. The stresses remain effectively constant throughout, and only plastic increments of strain need be considered.

For this example we can obtain the influence of boundary motion on the tensile force required for continued extension. The cross-sectional area of the block is reduced due to the lateral contraction $(-\epsilon_y)$, and this effect is also shown in fig. 3 as a percentage effect reducing the limit load below that based on the neglect of this influence. It is seen that this boundary influence increases with the strain and soon dominates the variation in surface traction as the asymptotic limit stress is approached. For this example the boundary motion effect only attains a magnitude of about $\frac{1}{2}\%$ for strains of the order 5 times the elastic limit strain, and so is itself quantitatively unimportant. However the possibility of an unstable necking contraction has not been considered which for long specimens would be likely to produce a much larger influence at such average strain magnitudes. The solution for compression in plane strain is essentially the same with reversed signs for the stress and strain components, so that the increase of the cross-section produces an increase in the limit load. However a complete analysis would have to include the influence of lateral instability, which would involve a larger boundary motion influence. Thus a study of this example indicates that the asymptotic approach to the limit stress may occur so rapidly that it is effectively reached for strains of elastic order, and the boundary motion may produce a larger influence on the limit load at this stage. If larger strains must be permitted before the analysis is valid, as in the case of approximating an actual body by a plastic-rigid model, the neglect of boundary motion may involve a

Fig. 3



serious error. For most problems, in which plastic flow occurs locally in regions adjacent to elastic regions in contrast to the uniform stress situation of this particular example, the redistribution of stress during contained flow is likely to anticipate the asymptotic approach to the limit stress, and so make it more rapid, while the concentration of boundary deformation in the active plastic regions is likely to produce a more marked influence of boundary motion. The latter is of course strongly emphasized in the case of slender bodies, and bodies of open section. There will, of course, be cases where the asymptotic approach to the limit load will be slower than for the particular example considered. Thus in cases where stress discontinuities occur in plastic-rigid analysis, which must be considered as the limit of narrow elastic regions (Lee 1950), slower approach to the limit is to be expected. The torsion example considered above for a solid shaft ($a=0$) would be such an example. However, in many cases effectively constant surface tractions occur at strains of elastic order while the influence of boundary motion is still small. The elastic-plastic limit load theorems will then provide good approximations to the conditions at which deflections begin to increase markedly without appreciable change of load, so that there is a well defined knee in the load deflection relation. In many cases thus satisfactorily analysed for an elastic-plastic material, the use of a rigid-plastic theory, which demands strains large compared with elastic strains before its predictions can reasonably be expected to comprise an approximation for an actual material, may not be valid due to the large boundary motion influence at such strains. It is for such cases that the application of the elastic-plastic theory is essential.

§3. UNIQUENESS THEOREMS

If we make use of the development leading to Theorem I of Drucker, Greenberg and Prager which shows that at the limit load for uncontained flow under constant surface tractions only increments in the plastic component of strain occur, the uniqueness and extremum principles proved by Hill (1951) for a plastic-rigid body can be shown to apply to an elastic-ideally plastic body at the limit load. The proofs carry through exactly as in the plastic-rigid case.

Thus if (σ_{ij}, u_i) and (σ^*_{ij}, u_i^*) are stress and velocity fields satisfying boundary conditions prescribing velocities over part of the surface and tractions over the rest, and each satisfying equilibrium and plasticity equations in its own regions of flow, and not violating the yield condition in regions of flow of the other solution, then σ_{ij} and σ^*_{ij} are identical throughout regions of non-zero strain rate of either u_i or u^*_i.

Also the principle of maximum plastic rate of work over parts of the boundary where velocities are prescribed and its complement apply for an elastic-plastic material at the limit load. This compares the rate of work for an actual solution, and that for the actual velocities and any equilibrium stress distribution σ^*_{ij} which does not violate the yield condition. In the case when tractions are prescribed over the whole surface, this implies that σ_{ij} for the actual solution is equal to σ^*_{ij} wherever deformation is occurring.

These principles for an elastic-plastic material at the limit load are much stronger than can be proved for an elastic-plastic material in the region of contained plastic flow.

The fact that at the limit load only plastic components of strain increment occur, enables plastic-rigid solutions to determine the stress distribution in the regions flowing plastically of an elastic-plastic body. The several examples which were presented (Drucker, Greenberg, Lee and Prager 1951) to demonstrate exact determination of limit loads, therefore also present the exact stress distribution in the plastic regions at the limit load for elastic-plastic materials. This conclusion provides a fuller means of assessment of the development of numerical solutions for contained flow.

In the proof of the theorems for a rigid-plastic material, the orthogonality between strain and stress increments is not needed so that the results go through formally for a work-hardening material. This is in contrast to the elastic-plastic analysis. However, since for application to an actual material, the rigid-plastic theory can only be expected to be satisfactory when strains large compared with elastic strains have occurred, the work-hardening due to these strains is neglected in the theory of the yield limit, and stress magnitudes will be uncertain to this extent. Thus the practical extent to which a work-hardening material is analysed is subject to this limitation, and within the uncertainty of such stress magnitudes, the elastic-plastic theory also applies, since prior work-hardening can be included as a variable yield limit. However, such an analysis cannot be considered a satisfactory treatment of the deformation of a work-hardening material.

APPENDIX

THE TENSION TEST IN PLANE STRAIN

We consider a block of elastic-plastic material stressed homogeneously in plane strain by the tensile stress σ_x, with $\epsilon_z=0$ and $\sigma_y=0$. Discussion of the stress strain response is facilitated by using the plane diagram obtained by projecting the three dimensional stress and strain spaces [plotted with the same coordinates $(\sigma_x, \sigma_y, \sigma_z)$ and $(2G\epsilon_x, 2G\epsilon_y, 2G\epsilon_z)$ where G is the elastic shear modulus] on the plane $\sigma_x+\sigma_y+\sigma_z=0$. Previously this has been considered as a plot of the stress and strain deviators along axes inclined at 120° to each other, but if we note that equal vectors along each axis produce a zero resultant, we can plot the stress and strain components directly and simplify the use of the diagram. Figure 4 shows this stress, strain plane, in which the Mises yield criterion for an ideally plastic material takes the form of a circle of radius Y, the yield stress in tension. For purely elastic strains, the stress and strain points coincide and lie within the yield circle. When plastic flow is occurring the stress point must be on the circle, while the strain point can move outside it.

It is convenient to specify the stress point during plastic flow by the angle θ which the stress vector makes with the horizontal, and the strain point by coordinates (ξ, η). Then for the particular problem under consideration we have :

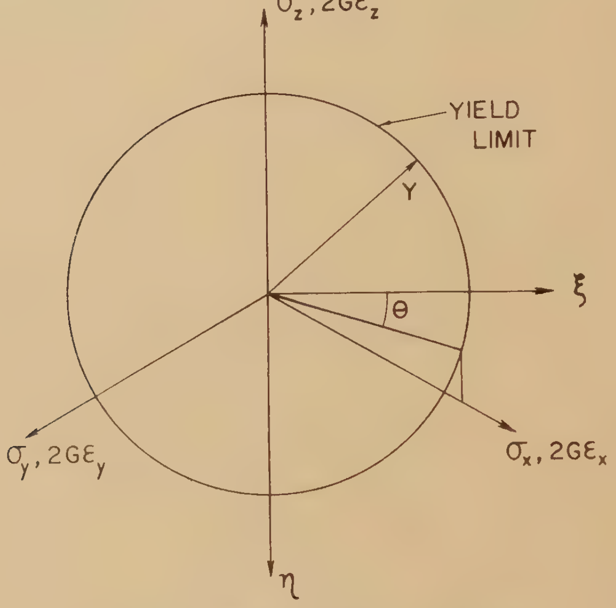
$$\sigma_x = 2Y/\sqrt{(3)}\cos\theta,\quad \sigma_z = Y/\sqrt{(3)}\cos\theta - Y\sin\theta, \quad \ldots \quad (1)$$

$$\xi = \sqrt{(3)}\,G(\epsilon_x - \epsilon_y),\quad \eta = G(\epsilon_x + \epsilon_y). \quad \ldots \quad (2)$$

The compressibility condition becomes

$$(3K/G)\,\eta = Y[\sqrt{(3)}\cos\theta - \sin\theta] \quad \ldots \quad (3)$$

Fig. 4


$\sigma_z, 2G\varepsilon_z$
YIELD LIMIT
Y
ξ
θ
$\sigma_y, 2G\varepsilon_y$
$\sigma_x, 2G\varepsilon_x$
η

The plane diagram for stress and strain.

and the Prandtl–Reuss relation connecting the total strain increment associated with an increment of stress :

$$\tan\theta = (d\eta - Y\cos\theta\,d\theta)/(d\xi + Y\sin\theta\,d\theta), \quad \ldots \quad (4)$$

(3) and (4) combine to give a directly integrable differential equation in ξ and θ, which in combination with the initial conditions based on the initial elastic strain determines :

$$\xi = Y\{\cos\theta_0 - \big|\log\tan(\theta/2) + (G/3K)[\sqrt{(3)}\sin\theta + \cos\theta + \log\tan(\theta/2)]\big|_{\theta_0}^{\theta}\} \quad \ldots \quad (5)$$

θ_0 is the direction of the stress vector during purely elastic deformation, which by use of (3) and the coincidence of the stress and strain points is given by

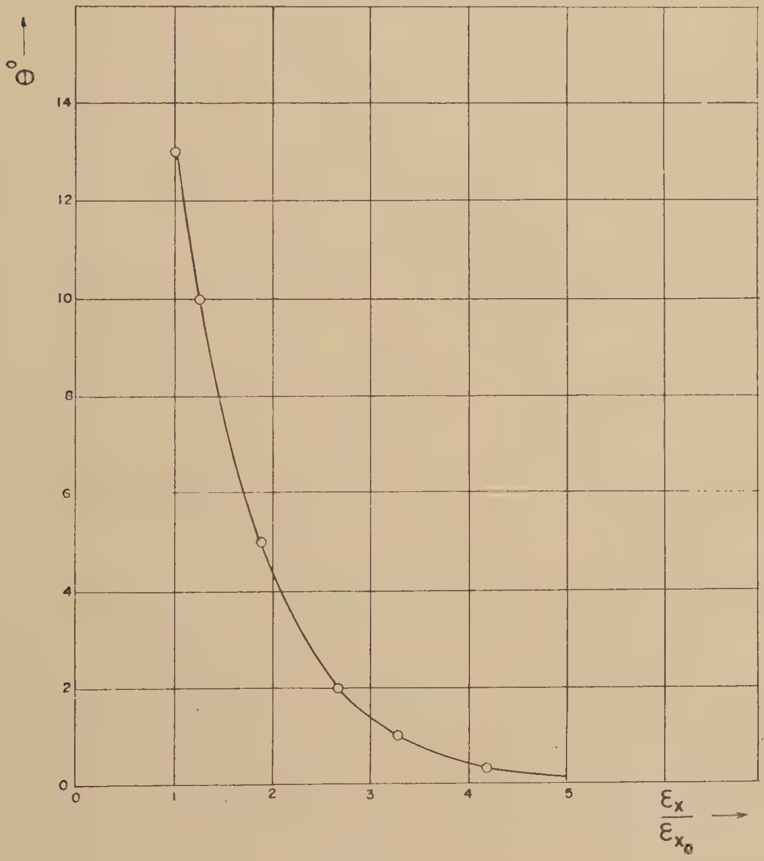
$$\tan\theta_0 = [\sqrt{(3)}/\{1 + (3K/G)\}]. \quad \ldots \quad (6)$$

In terms of (ξ, η) the strain components are given by

$$2G\epsilon_x=\{\eta+[\xi/\sqrt{(3)}]\}, \qquad 2G\epsilon_y=\{\eta-[\xi/\sqrt{(3)}]\}. \quad . \quad . \quad . \quad (7)$$

It is clear from fig. 4 that increasing strain ϵ_x corresponds to increasing ξ, and according to (5) this is associated with decreasing θ. In fact as θ approaches zero, ξ approaches infinity, so that as the extension proceeds θ decreases towards zero. In fig. 5 θ is plotted against the extension ϵ_x, expressed in units of the yield point extension, and this curve emphasizes the rapid decrease of θ towards zero. Thus from (1), σ_z approaches $Y/\sqrt{(3)}$ the disparity being of order θ, and σ_x approaches $2Y/\sqrt{(3)}$ with a disparity of order θ^2. This increase of σ_x and associated rapid approach to a limit is the asymptotic approach to the limit stress discussed in § 2 of the paper. The percentage disparity below the limit stress is plotted in fig. 3 against the extension. It is seen that the stress is within 0·1% of the limit stress when the strain is only $2\frac{1}{2}$ times the elastic limit strain.

Fig. 5



In order to obtain the variation of total load the increase in area normal to σ_x must be considered by computing ϵ_y from (7). This can be expressed as a percentage reduction of the tensile load below that based on the initial area. For this simple case, this represents the influence of the motion of the boundaries on the limit load, and is plotted in fig. 3 to provide a comparison with the influence of the asymptotic approach to the limit stress. It will be seen that for strains greater than three times the elastic limit strain, the reduction of area influence, although small, completely dominates the influence of the asymptotic approach to the limit stress.

REFERENCES

DRUCKER, D. C., GREENBERG, H. J., LEE, E. H., and PRAGER, W., 1951, *Brown University Technical Report No.* 61, to appear in *Proc. of* 1*st U.S. National Congress of Applied Mechanics.*

DRUCKER, D. C., GREENBERG, H. J., and PRAGER, W., 1951 a, *Jl. Appl. Mech.*, **18,** 371 ; 1951 b, *Brown University Technical Report No.* 59, *Quarterly of Applied Math.*, **9,** 381, 1952.

GREENBERG, H. J., and PRAGER, W., 1951, *Proc. Am. Soc. Civ. Engrs.* 77.

HILL, R., 1948, *Quarterly J. Appl. Math. and Mech.*, **1,** 18 ; 1950 a, *J. Appl. Mech.*, **17,** 64 ; 1950 b, *The Mathematical Theory of Plasticity* (Clarendon Press); 1951, *Phil. Mag.*, **42,** 868.

HILL, R., LEE, E. H., and TUPPER, S. J., 1947, *Proc. Roy. Soc.* A, **191,** 278.

LEE, E. H., 1950, *Proc. 3rd Symp. in Applied Math. of Amer. Math. Soc.*, 213.

PRAGER, W., HODGE, P. G., 1951, *Theory of Perfectly Plastic Solids.* John Wiley and Sons, Inc.

REUSS, A., 1930, *Z. angew. Math. Mech.*, **10,** 266.

Comments on Dr. Lee's Paper[*]

By R. HILL
Department of Theoretical Mechanics, University of Bristol

Dr. Lee claims that the so-called plastic-elastic limit theorems of Prager and co-workers are essentially different from my theorems relating to the yield point of a plastic-rigid body, and that they have a greater practical value. My view is that the theorems are formally identical and give exactly the same results.

The stress in a *p-r* body under specified surface tractions is obtained by indefinitely increasing the rigidity modulus G in a *p-e* solution. The influence of G is likely to be negligible whenever the plastic strain exceeds the elastic strain by a factor of order 3 (the straining being reasonably monotonic). However, these are by no means the only circumstances (as Dr. Lee maintains) where *p-r* analysis gives useful results. It has been proved, for instance, that in a wide class of plane strain problems the stresses are independent of G, even when plastic strains are of elastic order (Hill 1950 b, p. 244). This is true also in a bar under given torque or in a hollow sphere under given pressure (Hill 1950 b, pp. 87 and 99).

---

[*] With the consent of the author, Dr. Lee's paper was shown by the editors to Dr. Hill.

A precise meaning is attached to the limit or yield-point load of a *p-r* body: it corresponds to the moment when local distortion *first* begins under the given loading programme. By contrast, the load-distortion curve for a *p-e* body usually ascends smoothly from zero, and only when it passes through or attains a stationary value can a precise meaning be attached to the term *p-e* yield point. If changes of geometry and hardness are neglected in the *p-e* analysis, then, under reasonably monotonic loading, the load-distortion curve approaches the *p-r* yield point load asymptotically, though it does not usually reach it in any finite strain (e.g. Hill and Siebel, 1951, *Phil. Mag.*, **42,** 722). (The *p-r* work-rate is approached from *below*, according to the maximum work principle.) When geometry changes are not neglected, the *p-e* curve has a more or less pronounced knee to which the *p-r* yield point roughly corresponds. The *p-r* yield point for the *undeformed* body has, therefore, a clear practical significance. Dr. Lee perversely maintains, however, that the *p-r* analysis must be based on the *deformed* body after some unspecified strain, while allowing that the *p-e* analysis based on the *undeformed* body is satisfactory.

Consider, next, the maximum work principle and its complement (Hill 1950 a and 1950 b, p. 66). The associated continued inequality sets arbitrarily close bounds to the work-rate at the yield point of a *p-r* body. This is analytically identical with that afterwards given by Drucker, Greenberg and Prager (1951), and Prager and Hodge (1951, equations 37.23, 39.12, 39.14). This later work, being restricted to proportional loading of non-hardening material, is on these two counts a special case of my 1950 theorems; the results of Greenberg and Prager (1951) for a structure of hinged bars are still more specialized. On the other hand, this later work was more general than mine in that it was formulated for a partly-plastic body; examples of its use were also given.

However, Prager stated, without reservation (e.g. Prager and Hodge 1951), that the continued inequality provides arbitrarily close bounds to the *p-e* limit load, but only proved this *if* deformation occurs under constant load when geometry changes are neglected. As already remarked, such a stationary load is usually approached at infinite strain and then only if the straining is reasonably monotonic; if boundary motion is allowed for, the curve rises steadily when the system is stable. Thus, Prager's interpretation is approximate to a degree that is impossible to state beforehand. The only *precise* meaning that can *invariably* be given to the load concerned in these theorems is that it is the yield point of a *p-r* body. Accordingly I disagree entirely with Dr. Lee's opinion that 'the *p-e* approach provides definite information, whereas nothing can be deduced with certainty from the *p-r* theory'. Elsewhere (Drucker, Greenberg, Lee and Prager 1951) Dr. Lee seemingly contradicts himself: 'the application of a complete *p-r* solution in the theorems gives the exact limit load for the *p-e* problem'. Furthermore, one of the illustrations cited was one where the load rises steadily (compression of a block between rough plates).

LI. *Growth Spirals Originating from Screw Dislocations on Gold Crystals*

By S. Amelinckx
Geological Institute, University of Ghent*

[Received January 11, 1952, revised MS. January 28, 1952]

Summary

The present paper gives a more detailed account of the observation of growth spirals on (111) faces of gold crystals, obtained by precipitation. The observations prove that Frank's theory also holds for metal crystals; moreover they constitute the first visual proof for the presence in metal crystals of dislocations in general, and screw dislocations in particular. They further show that the Burgers vector of dislocations in metals is *not* necessarily limited to *one* unit of a few simple lattice displacements, as is generally assumed. The shape of the growth spirals is briefly discussed.

§1. Introduction

According to the theory of Burton, Cabrera and Frank (1949, 1951) crystals can only grow, at supersaturations too low for two-dimensional nucleation, when dislocations are present. Where a screw dislocation ends in a crystal face an exposed edge is formed which, whilst growing, will subsist and wind itself up in the shape of a spiral (Frank 1949). Recently several experimental confirmations of this theory have been found on crystals of beryl (Griffin 1950), n-hexatriacontane (Dawson and Vand 1951), carborundum (Amelinckx 1951 a, c; Verma 1951 a, b), and cadmium iodide (Forty 1951). In the case of mica the presence of dislocation lines with a screw component normal to the cleavage face has been proved by means of multiple beam interferometry (Amelinckx, in the press). Arguments strongly suggest that this is also the case for other layer lattice minerals such as kaolinite, nacrite, dickite (Dekeyser and Amelinckx 1951).

Although the concept of a dislocation has been introduced to explain the mechanical properties of metals, no direct experimental evidence has been obtained so far for their existence in metals. The present paper gives a more detailed account of the observation of growth spirals on gold crystals, obtained by precipitation, and offers as a consequence an almost direct visual proof for the presence of screw dislocations in metal crystals. A preliminary account of these observations, in which the preparation of the gold crystals in question was described, has been published elsewhere (Amelinckx, Grosjean and Dekeyser, 1952).

---

* Communicated by Professor W. Dekeyser.

§2. DESCRIPTION OF CRYSTALS

The crystals were tabular ; they measured 5–40 microns across, and had hexagonal or trigonal shape. In fact they were octahedra developed in a tabular way following (111). Gold, having a face centred cubic lattice (a=4.07 Å), can be considered as a close packing of spheres. The (111) planes are the close-packed and consequently the normal habit faces ; at low supersaturations such faces can only grow by means of spiral growth. As described in detail elsewhere (*loc. cit.*), the crystals were obtained by boiling a standard gold plating solution with concentrated hydrochloric acid. When this mixture had cooled, crystals of sodium chloride and of gold were found on the bottom of the vessel, after a few days. It was therefore suggested by Prof. De Block (private communication), that the gold crystals had probably grown by epitaxy. It is indeed known (Van der Merwe 1949) that gold crystallizes with its (111) face on (110) of NaCl. The importance of this remark will appear afterwards (§ 6).

§3. TECHNIQUE OF OBSERVATION

The crystal faces were observed by means of an ordinary metallurgical microscope using bright field illumination. The photographs were taken slightly off focus, as this procedure increases the visibility of small surface steps. It has been shown that steps of only 7 Å high are still visible by this method, when conditions are good (Amelinckx 1951 b). As most of the crystals had their (111) face not exactly normal to the incident light beam, conditions were not favourable for observation. As a consequence only spirals with a relatively great step height could be photographed.

Owing to the small dimensions of the crystals it was necessary to use the highest magnification. The consequent short working distance of the objective prohibited, unfortunately, all conventional interferometric measurements. The following procedure however enabled an estimate of the heights of steps. A small drop of a water-alcohol mixture was put on the crystal face. After a few seconds this drop evaporates partially, leaving a thin layer of liquid. Between the interfaces liquid-crystal and liquid-air two-beam interference fringes of the Fizeau type are formed, when using parallel, monochromatic light at normal incidence. Greatest dispersion is obtained just before complete evaporation. Unfortunately this procedure only allows visual interpretation as the fringes are constantly moving.

§4. SHAPE AND ORIENTATION OF GROWTH SPIRALS

On approximately 10% of the gold crystals investigated, growth spirals with a well defined centre have been observed on the octahedral faces ; typical photographs are shown on plate XXIV.*

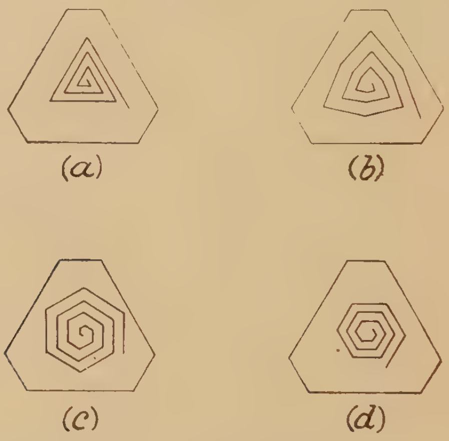
In comparison with the growth spirals visible on carborundum their development is poor. In most cases only a few turns can be seen. The spiral growth fronts are not smooth but 'kinky' (figs. 2, 3, pl. XXIV). The degree of 'polygonization' (Amelinckx 1951 c) is always very near to unity.

---

* For plate see end of issue.

The shape of the spirals is dependent upon the rate of advance of a growth front in different crystallographic directions. It will therefore be triangular as a consequence of the triad symmetry axis perpendicular to (111) in cubic crystals. The loops of the spiral are parallel to the edges of the octahedral face (fig. 8 (*a*) and figs. 1, 2, pl. XXIV). There are thus three main minima for the radial growth velocity corresponding to the directions perpendicular to the edges of the octahedral faces marked (*a*), (*b*) and (*c*) on fig. 9. These directions are the most close-packed.

Fig. 8

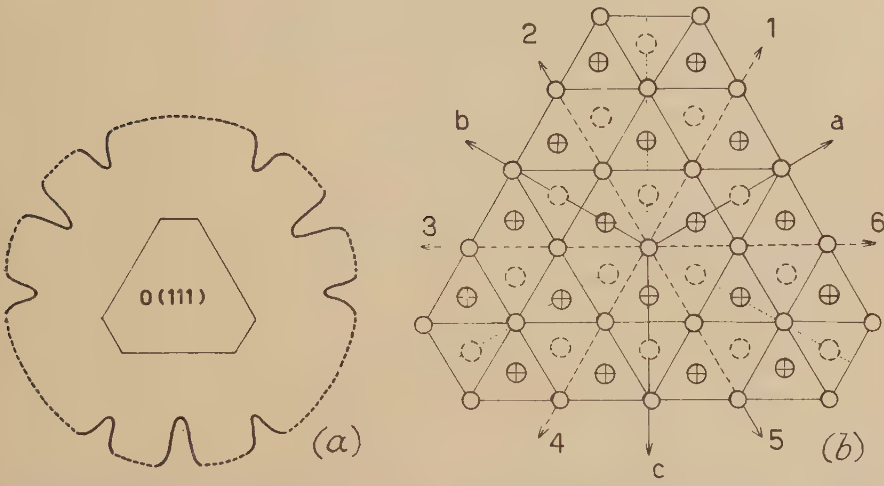


Idealized shape and orientation of growth spirals.
- (*a*) triangular spiral, main minima only determine the shape.
- (*b*) intermediate shape: main and secondary minima are determining the spiral's shape.
- (*c*) hexagonal spiral: the secondary minima dominate the main minima.
- (*d*) probable orientation of growth spiral on hypothetical hexagonally close-packed gold.

There are six secondary minima (1, 2, . . 6). In certain circumstances, which are not clear for the moment, these secondary minima influence the shape of the spiral (figs. 5, 6, pl. XXIV). The symmetry remains trigonal: the idealized shape becomes as represented on fig. 8 (*b*). In the extreme case when the influence of the secondary minima is dominating, the shape becomes hexagonal (fig. 8 (*c*) and fig. 7, pl. XXIV); the hexagonal spiral being rotated 30° with respect to the hexagonal outline of the crystal. Polar diagrams for the radial growth velocity, as deduced from the shape of the spirals, are represented on fig. 10 for the two extreme cases (*a*) for fig. 1, pl. XXIV, and (*b*) for fig. 7, pl. XXIV.

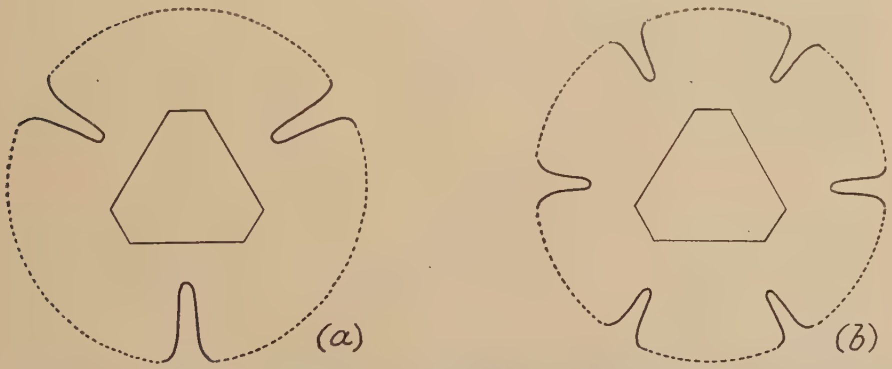
In rare cases hexagonal spirals with their branches parallel to the face's edges were observed (fig. 4, pl. XXIV). It is suggested that these could be hexagonally close-packed crystals, making the symmetry on (111) really hexagonal. It is indeed sufficient that the sequence *ABAB* . . . is formed once, by mistake, and that a suitable dislocation is present, to make sure that this arrangement will be repeated throughout the crystal or a part of it (Amelinckx 1951 c). This could however not be checked, the crystals being too small for single crystal x-ray diffraction. A powder diagram of the bulk material showed only the lines of the cubic lattice.

Fig. 9



Crystal structure of gold seen along [111]. The directions of main (*a*, *b* and *c*) and secondary (1, 2, 3, . . 6) minima in growth velocity are shown with respect to the lattice and to the outline of the octahedral face.

Fig. 10



Polar diagrams for the radial growth velocity as a function of the crystallographic direction in the (111) plane, for the spirals of fig. 1 (*a*) and fig. 7 (*b*) of plate XXIV.

The kind of mistake mentioned above is probably one of the origins of dislocations.

To estimate the heights of the steps the interferometric procedure described above was used. The deviation of the Fizeau-fringes caused by a known number of steps was estimated. In one case we found a step height of the order of 400 Å ± 100 Å (fig. 6, pl. XXIV). In other cases no deviation of the fringe-system could be seen (figs. 1, 5, pl. XXIV) proving that the steps are less than 100 Å.

These measurements prove that the spirals cannot be due to dislocations of unit strength ($d_{111}=2, 2$ Å), but must be generated by groups of dislocations all of one sign, having resultant Burgers vectors with a screw component, normal to (111), of a rather great number of repeat distances long.

§5. TOPOLOGICAL CONSIDERATIONS

In view of the calculation of the normal growth velocity (velocity normal to the crystal face) it is necessary to have information about the topological relations between dislocations, when the spirals generated by them are interacting. In the case of gold we only observed single or multiple growth promoting dislocations generating only *one* spiral dominating the growth of the face. Left- and right-handed spirals were seen in approximately equal numbers.

No spirals of opposite sign, interacting so as to form closed terraces, were seen, although special attention was paid to this and a very large number of crystals were investigated. A possible explanation is as follows. Dislocations of opposite sign attract each other and try to destroy one another, the attraction being roughly proportional to the product of the strengths of the dislocations and to the inverse of the distance. In the case of carborundum dislocations of opposite sign, situated very near one to another, can be observed ; in metals however the mobility of dislocations is much greater, so that it is probable that dislocations of opposite sign and of suitable strength to produce visible spirals, cannot coexist within the small growing nucleus.

§6. DISLOCATION TUBES

Very recently Frank (1951) has shown theoretically that when the Burgers vector of a screw dislocation exceeds a certain critical value, (10 Å), the dislocation line is the axis of a hollow tube, in the interior of the crystal. The end of this tube has the shape of a crater. Frank considers the dots seen in the centres of growth spirals of carborundum as being caused by these craters, thus confirming his theory. He states however that these craters are not likely to occur in metals. It is now clearly visible on our photographs that the point of emergence of the dislocation line is marked by a dot. Apparently a discrepancy exists between theoretical expectation and the interpretation of observations. The reason for this is simply that Frank did not consider the possibility of dislocations having a Burgers vector exceeding 10 Å in metal crystals. Dislocations of the value mentioned above (400 Å) are indeed unlikely to occur in metal crystals, unless they have grown in very special circumstances. Such circumstances are for instance, according to Frank (private communication), growth on a foreign crystallization nucleus, in which case the gold crystal would inherit the screw dislocations present in the host crystal (this is in accordance with the growth process postulated in § 2). Once the large hollow screw dislocations are present, they are stable, as a large activation energy is necessary to detach unit dislocations from them.

It seems to us that the large amount of misfit (25%) between the gold lattice and the host lattice is also responsible, at least partly, for the creation of dislocations.

ACKNOWLEDGMENTS

I wish to thank Prof. Dr. W. Dekeyser for continuous interest and useful discussion, and Dr. C. Grosjean for preparing the gold crystals. To Dr. F. C. Frank I am much indebted for advice. I gratefully acknowledge support from the 'Centre National Belge de Chimie–Physique Moléculaire'.

REFERENCES

AMELINCKX, S., 1951 a, *Nature, Lond.*, **167,** 939; 1951 b, *Ibid.*, **168,** 431; 1951 c, *Journ. de Chim. Phys*, **48,** 425.
AMELINCKX, S., to be published in *Nature, Lond.*
AMELINCKX, S., GROSJEAN, C. and DEKEYSER, W., 1952, *Comptes Rendus*, **234,** 113.
BURTON, W. K., CABRERA, N., and FRANK, F. C., 1949, *Nature, Lond.*, **163,** 398; 1951, *Phil. Trans. Roy. Soc.* A, **243,** 299.
DAWSON, I. M., and VAND, V., 1951, *Nature, Lond.*, **167,** 476.
DEKEYSER, W. and AMELINCKX, S., 1951, *Comptes Rendus*, **233,** 1297.
FORTY, A. J., 1951, *Phil. Mag.* [7], **41,** 760.
FRANK, F. C., 1949, *Discussions of the Faraday Society, No.* 5, **48,** 67; 1951, *Acta Cryst.* **4,** 497.
GRIFFIN, L. J., 1950, *Phil. Mag.* [7], **41,** 196.
VAN DER MERWE, J. H., 1949, *Discussions of the Faraday Society, No.* 5, 212.
VERMA, A. R., 1951, *Nature, Lond.*, **167,** 939; *Phil. Mag.* [7], **42,** 1005.

LII. *Experiments on Two-dimensional Supersonic Flow in Corners and over Concave Surfaces*

By N. H. Johannesen
Fluid Motion Laboratory, University of Manchester*

[Received February 4, 1952]

Summary

Two simple cases of two-dimensional supersonic flow were investigated by schlieren photography, viz. the flow in two consecutive corners and the flow over a circular-arc profile concave to the stream. The flows were photographed for smooth and rough surfaces of the models, i.e. with laminar and turbulent boundary layers, and the observed flow patterns are compared with those predicted by inviscid theory. The flow with a turbulent boundary layer was in good agreement with inviscid theory outside the boundary layer, whereas the laminar boundary layer separated and a shock wave originated at the point of separation. In both cases the agreement was good at points far from the surface.

A discussion is given of the flow at the point of intersection of three shock waves, and of a shock wave and a centred compression. For the latter case photographs show a weak shock wave in the direction of the downstream Mach line at the point of intersection.

§1. Introduction

When a uniform two-dimensional supersonic stream meets two consecutive sharp corners, inviscid theory predicts two shock waves originating at the corners and combining into one shock wave at some distance from the surface. Further, since in general this simple three-shock configuration is not by itself a possible solution of the flow equations, it is necessary to introduce a weak expansion or shock wave at the point of intersection.

When the uniform stream meets a circular-arc profile which is concave to the flow and has its initial direction parallel to the free stream, inviscid theory predicts a shock-free flow (compression in a simple wave) near the surface, with a shock wave starting on the leading Mach line at some distance from the surface.

These two cases of inviscid flow are considered in § 3 below. The discussion includes a treatment of the flow near the point at which a vortex sheet originates as the result of the interaction of three shock-waves or of a shock wave and a centred compression. For the latter case the sign of the weak disturbance (compression or expansion) at the point of intersection is determined. The flow over the circular-arc profile is

---

* Communicated by W. A. Mair.

examined by combining the characteristic equations for rotational flow and the shock-wave equations. A numerical example is computed for a free-stream Mach number of 1.97.

In § 4 the experiments are described and the results compared with the inviscid flows. The flows were photographed for smooth and rough surfaces of the models, i.e. with laminar and turbulent boundary layers. The upstream effect of the shock wave or compression was found to be different in the two cases. In some cases the weak disturbance mentioned above could be detected.

§2. NOTATION

M	Mach number.
P	ratio of local pressure to free-stream pressure.
R	gas constant.
a	velocity of sound.
p	pressure.
q	velocity.
r	radius of circular-arc profile.
s	entropy.
t	function defined in eqn. (6).
x, y	Cartesian coordinates.
γ	ratio of specific heats.
δ_1, δ_2	deflections at first and second corners.
θ	direction of stream.
μ	Mach angle.
ρ	density.

$$\sigma = \frac{s}{2\gamma R}.$$

Suffix 1 denotes free-stream conditions. Other suffixes are defined in the figures.

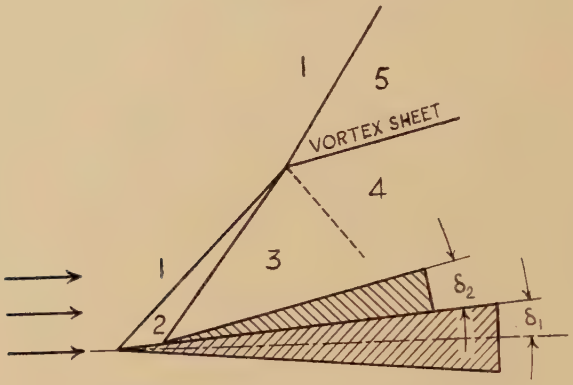
§3. THE INVISCID FLOWS

We assume the free stream to be uniform and the flow over the obstacles to be adiabatic. The entropy is assumed to be constant along each streamline except through a shock wave. The effects of viscosity, and thus the boundary layers, are ignored. We deal with the two cases separately.

(a) *Flow in two consecutive corners.* Let us consider the flow over a profile consisting of a small wedge placed on top of a larger wedge (fig. 1), and let the deflections be δ_1 and δ_2. The two shock waves in the corners are geometrically fixed, and the conditions in region 3 depend on the free-stream conditions and on δ_1 and δ_2. We shall assume that the flow in 3 is supersonic.

The two shock waves combine into one, and a vortex sheet originates at the point of intersection. Across the vortex sheet the pressure and flow direction must be continuous. These conditions can, in general, not be satisfied by such a three-shock configuration, and it is necessary to introduce an additional weak shock wave or expansion at the point of intersection. This weak disturbance must lie between the shock wave (2–3) and the vortex sheet, separating the regions 3 and 4. Its direction is approximately that of a Mach line.

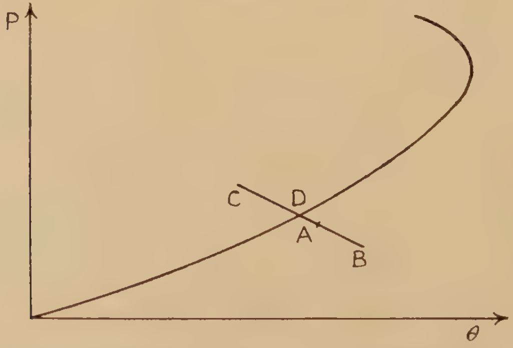
Fig. 1



Flow over double wedge.

Let p denote the pressure, P the ratio of pressure to free-stream pressure p_1, θ the velocity direction measured from the free-stream direction, and M the Mach number. For the single oblique shock wave (1–5) P_5 can be plotted against θ_5 for a fixed M_1 (fig. 2). The point

Fig. 2



Relations between pressure ratios and streamline deflections.

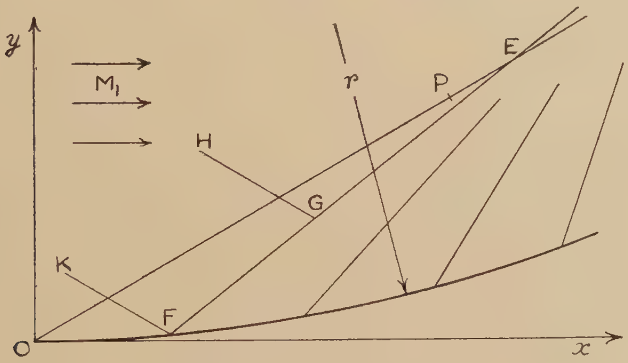
A with coordinates (θ_3, P_3) will, in general, not lie on the curve $P_5=P_5(\theta_5)$. The points (θ_4, P_4) for various strengths of the weak disturbance (3–4) will then lie on the curve BAC, where the branch AB corresponds to an expansion and AC to a compression. The position of A in relation to the curve $P_5=P_5(\theta_5)$ determines whether the disturbance is an expansion or a compression. The small deflection at the disturbance is $\theta_D-\theta_A$, where D is the point of intersection of the two curves.

For the values of M_1, δ_1 and δ_2 of the experiments described below the disturbance was found to be a compression, but its strength was so small that the shock-wave could not be seen on the photographs. Thus for $M_1=2$, $\delta_1=5°$ and $\delta_2=10°$, $\theta_D-\theta_A=-0{\cdot}02°$ and $(p_4-p_3)/p_3=0{\cdot}001$. This is only about a tenth of the minimum strength that can be detected by the schlieren system.

The theory of the three-shock configuration was not investigated in detail, but the simpler case where the two corner shock-waves are replaced by a continuous centred compression is considered in the following section.

(*b*) *Flow over a concave surface.* Assume a uniform stream of Mach number M_1 meeting a circular-arc profile which is concave to the flow and has its initial direction parallel to the free stream (fig. 3). Near the surface the compression takes place through a simple wave, but the Mach lines from the surface will converge and intersect, and a shock wave will be formed at some distance from the surface.

Fig. 3



Flow over concave surface.

The flow pattern can be computed by combining the characteristic equations for rotational flow and the shock-wave equations. In two-dimensional rotational flow there are three families of characteristics: the plus and minus Mach lines* and the streamlines.

With Cartesian coordinates x, y the characteristic equations for the Mach lines can be written (see Meyer (1948 a), pp. 205–206)

$$\frac{dy}{dx}=\tan(\theta\mp\mu), \quad (1)$$

$$d\theta\mp\frac{\sin\mu\cos\mu}{\rho a^2}dp=0, \quad (2)$$

---

* An observer looking downstream from a given point has the plus Mach line through the point on his right and the minus Mach line on his left.

where the upper and lower signs refer to a plus and a minus Mach line respectively. θ is the angle between the x-axis and the velocity direction, μ the Mach angle, ρ the density, p the pressure and a the local velocity of sound.

For the streamlines we have

$$\frac{dy}{dx}=\tan\theta, \quad ds=0, \qquad (3)$$

$$q\,dq+\frac{a^2 d\rho}{\rho}=0, \qquad (4)$$

where s is the entropy and q the velocity magnitude.

Equation (2) can be written in the form

$$d(t\pm\theta)+d\sigma\sin 2\mu=0 \qquad (5)$$

with $dt=(dq/q)\cot\mu$ and $d\sigma=(ds)/(2\gamma R)$, where γ is the ratio of the specific heats and R the gas constant.

It is found that

$$t=\mu-\pi/2+\lambda^{-1}\cot^{-1}(\lambda^{-1}\tan\mu) \qquad (6)$$

with $\lambda=[(\gamma-1)/(\gamma+1)]^{1/2}$. The function $(1\,000-t)$, with t in degrees, has been tabulated (see Herbert and Older (1946)).

The shock-wave equations are most conveniently written so that the conditions downstream of the shock are expressed in terms of the shock-wave angle and the conditions upstream of the shock wave (e.g. see Ferri (1949), p. 45).

When the flow is isentropic, the streamlines are no longer characteristics. The equations of the Mach lines become

$$d(t\pm\theta)=0, \quad \frac{dy}{dx}=\tan(\theta\mp\mu). \qquad (7)$$

Near the curved surface the flow is isentropic. In fig. 3 let OE be the leading Mach line, FGE a minus Mach line, and KF and HG the plus Mach lines through F and G. Using (7) we find

$$\left.\begin{aligned} t_1+\theta_1&=t_F+\theta_F=t_G+\theta_G,\\ t_G-\theta_G&=t_F-\theta_F,\end{aligned}\right\} \qquad (8)$$

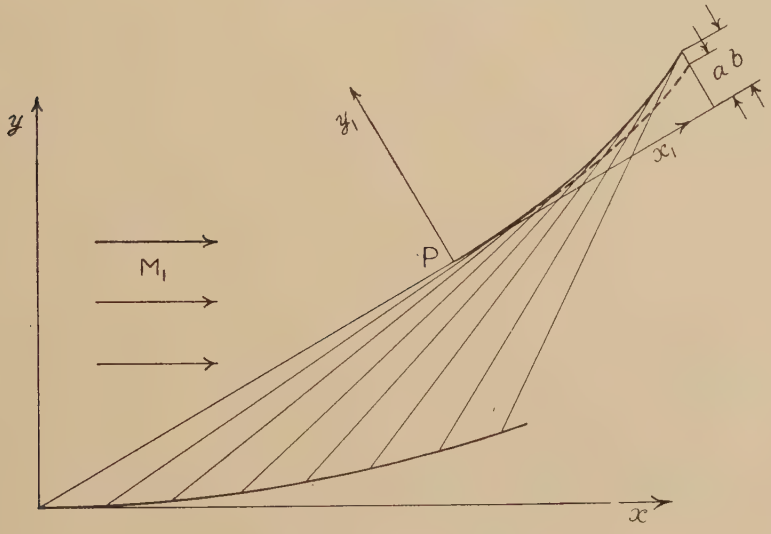
or $\theta_F=\theta_G$ and $t_F=t_G$, i.e. $\mu_F=\mu_G$, which shows that near the wall the minus Mach lines are straight. These Mach lines form an envelope as shown in fig. 4. The envelope starts at a point P on the leading Mach line. The position of P is found as the limiting position of E when the Mach line FE approaches the leading Mach line OE. It is found that

$$y_P=\frac{2r}{\gamma+1}\cdot\frac{M_1^2-1}{M_1^4}, \qquad (9)$$

where r is the radius of the circular arc. This result agrees with the condition for shock-free flow at the inlet of a two-dimensional contraction, derived by Meyer (1948 b, p. 455).

The envelope is situated upstream of the leading Mach line, which may be considered the second branch of the envelope (see Courant and Friedrichs (1948), p. 115). In order to obtain a simple approximate method of predicting the shock position, the ratio *a*/*b*, fig. 4, defining the position of the shock wave (shown as a broken line) in relation to the two branches of the envelope, was determined under the following assumptions: (i) the Mach lines are straight, (ii) the slope of the shock wave is the mean of the slopes of the Mach lines upstream and downstream of the shock wave, and (iii) the shock wave has the shape $y_1 \propto x_1^\alpha$ (see the x_1, y_1 system in fig. 4). The ratio *a*/*b* was found to be 0·75 for $\alpha=2$, 0·71 for $\alpha=3/2$, and 0·78 for $\alpha=3$. The value of 0·75 would therefore be expected to be nearly correct under widely different conditions.

Fig. 4



Envelope of Mach lines.

The flow behind the shock wave is rotational, but the entropy is constant along each streamline, and the flow can therefore be computed by using eqns. (3)–(6). In the general case of a supersonic rotational flow, it is necessary to cover the flow field with a network of Mach lines and streamlines and compute the properties of the flow and the geometry of the network simultaneously (see Ferri (1949), p. 86). However, in our case it is possible to use an easier procedure by which the properties at all the points in the network can be determined without knowing the positions of these points, i.e. the geometry. The geometry can then be determined afterwards by simple trigonometry.

In fig. 5 let *PQR* be the shock wave, *RLMSN* a minus Mach line, *PS* the plus Mach line through the point *P* (at which the shock wave starts), *PM* the streamline through *P*, and *QL* a streamline between *P* and *R*.

We then have from (5) with suffix R denoting conditions *downstream* of the shock wave at R

$$(t_R - \theta_R) - (t_S - \theta_S) + [d\sigma \sin 2\mu]_S^R = 0 \qquad . \; . \; . \; . \; . \qquad (10)$$

and from (8)

$$\theta_S = \theta_N \; ; \; t_S = t_N \; ; \; \mu_S = \mu_N \; ; \; t_S + \theta_S = t_1 . \qquad . \; . \; . \; . \; . \qquad (11)$$

Let the free-stream values s_1 and σ_1 be zero. As a first approximation we then have

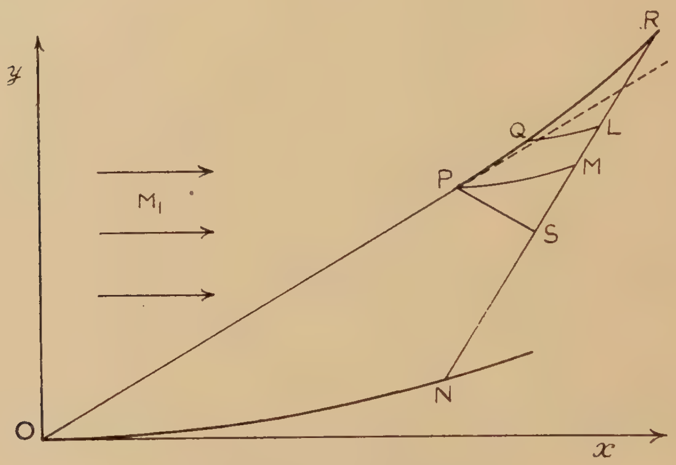
$$[d\sigma \sin 2\mu]_S^R = \sigma_R \sin (\mu_R + \mu_S)$$

and therefore

$$t_R - \theta_R - 2t_S + t_1 + \sigma_R \sin (\mu_R + \mu_S) = 0. \qquad . \; . \; . \; . \; . \qquad (12)$$

For known conditions at R, μ_S and t_S can be determined by trial and error, using (12) and (6). θ_S is found from (11).

Fig. 5



Computation of flow pattern.

For any point L on the minus Mach line between R and S we have

$$(t_R - \theta_R) - (t_L - \theta_L) + [d\sigma \sin 2\mu]_L^R = 0,$$

$$(t_L - \theta_L) - (t_S - \theta_S) + [d\sigma \sin 2\mu]_S^L = 0$$

or

$$(t_R - \theta_R) - (t_S - \theta_S) + [d\sigma \sin 2\mu]_L^R + [d\sigma \sin 2\mu]_S^L = 0 \qquad . \; . \qquad (13)$$

and, approximately

$$[d\sigma \sin 2\mu]_L^R = (\sigma_R - \sigma_Q) \sin (\mu_R + \mu_L),$$

$$[d\sigma \sin 2\mu]_S^L = \sigma_Q \sin (\mu_L + \mu_S).$$

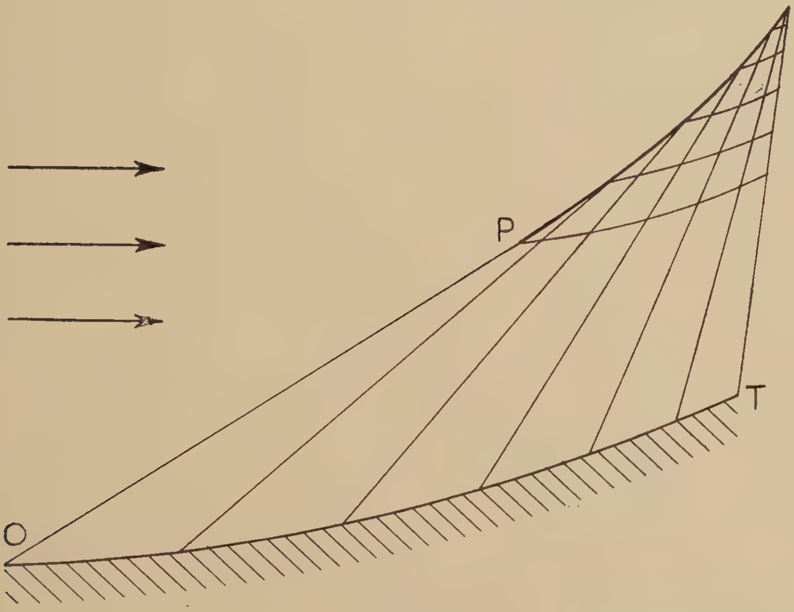
Eqn. (13) gives μ_L by trial and error, and the remaining conditions at L can then be found.

When sin $(\mu_R+\mu_L)$ and sin $(\mu_L+\mu_S)$ are nearly equal to unity, μ_L cannot be determined accurately from eqn. (13), but must be found by interpolation from the values at neighbouring points.

Fig. 6 shows the flow that was computed for $M_1=1{\cdot}97$, $\gamma=7/5$ and a total stream-deflection of 22·5° at the curved surface. (At this value of M_1, the velocity would be sonic after a deflection of 25·5°.) The ratio a/b as defined above was found to be 0·75 to 0·80, in good agreement with the simple estimate.

Because of the big steps used in the computations, fig. 6 could only be expected to show the general character of the flow pattern predicted by inviscid theory. However, the construction of the flow pattern was continued graphically, assuming that there is an expansion region with centre at T (fig. 6), and when the position of the decaying shock wave was compared with the corresponding experiments, the agreement was found to be good (see below).

Fig. 6



Flow over circular-arc profile for $M_1=1{\cdot}97$.

In order to investigate the possibility of a flow with a detectable weak shock wave at the point where a vortex sheet starts, we consider the case where the curved surface is so shaped that the continuous compression is centred and the envelope reduced to a point (fig. 7). Such a flow is physically possible, and can be obtained by giving the fixed boundary the shape of a streamline in a suitable Prandtl–Meyer expansion.

With the notation of fig. 7, we find the curves $P_5=(p_5/p_1)=P_5(\theta_5)$ and $P_3=P_3(\theta_3)$ similar to the curve $P_5=P_5(\theta_5)$ in fig. 2. If we expand the pressure ratios P_3 and P_5 in powers of the deflection θ, we get :

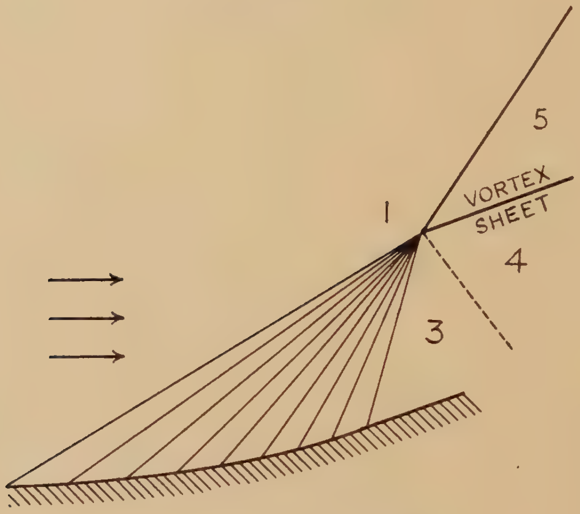
$$\left.\begin{aligned} P_3&=1+C_1\theta+C_2\theta^2+C_3\theta^3+0(\theta^4),\\ P_5&=1+C_1\theta+C_2\theta^2+(C_3-D)\theta^3+0(\theta^4),\end{aligned}\right\} \quad . \ . \ (14)$$

where

$$D=\frac{\gamma+1}{96}\gamma M_1^{6}(M_1^{2}-1)^{-7/2}[(5-3\gamma)M_1^{4}-(12-4\gamma)M_1^{2}+8] \quad . \ (15)$$

(see Ferri (1949), p. 76, and Kahane and Lees (1948)*). $D=0$ for $M_1=1{\cdot}245$ and $M_1=2{\cdot}540$ and is negative between these two values of M_1. Therefore, for small values of θ, the weak disturbance is a compression for $1{\cdot}245<M_1<2{\cdot}540$ and an expansion for $M_1<1{\cdot}245$ or $M_1>2{\cdot}540$.

Fig. 7



Flow with centred compression.

The exact values of M_1 for which $(P_5-P_3)=0$ depend on θ, and for high values of θ they are slightly different from those given above.

For the case of $M_1=1{\cdot}97$ and a deflection of 22·5° through the continuous compression (corresponding to the experiments described below) the weak shock wave was found to have a strength $(p_4-p_3)/p_3=0{\cdot}03$, and we should therefore expect to be able to see it on the schlieren photographs.

---

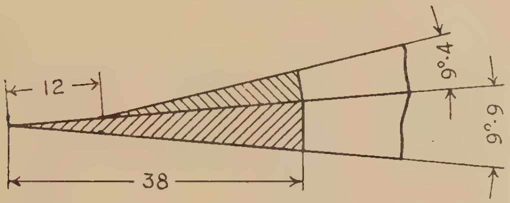
* It should be noted that eqn. (15) is given incorrectly in Busemann's original paper (*Lufo.* **12,** 210–220 (1935)) and in a number of other papers and books.

§4. EXPERIMENTS

The experiments were made in a small intermittent supersonic wind tunnel which has been described by Bardsley and Mair (1951). The working section of this tunnel is 12·5 × 10 cm, the Mach number 1·96 and the Reynolds number $1{\cdot}3 \times 10^5$ per cm. The humidity of the air was in all cases less than one part in 2 000 by weight and usually much lower. The models were made of high-carbon steel, ground and lapped to the correct shape. They extended over the whole width of the tunnel, and were supported by two arms on the downstream side. The position of these arms could be altered to give different angles of incidence of the models. The schlieren system was of the conventional two-mirror type, and a flash tube was used as light source because it was conveniently available. The knife edges were in all cases parallel to the free stream. The flows were also observed on a screen using a continuous light source, and were found to be steady.

(*a*) *The double wedge* had the dimensions shown in fig. 8. Experiments with smooth surfaces were made for δ_1 approximately 0, $2\frac{1}{2}$, 5, $7\frac{1}{2}$ and 10 degrees. Figures 9 and 10 (Plate XXV*) show the flows for $\delta_1 = -0{\cdot}1°$ and $\delta_1 = 7{\cdot}2°$. δ_2 was 9·4° in both cases. For $\delta_1 = 10°$ the tunnel was choked.

Fig. 8



Section of double wedge (with dimensions in mm).

In figs. 9 and 10 (Plate XXV) it can be seen that, in addition to the shock waves at the leading edge and in the corner, there is a shock wave from the point where the laminar boundary layer separates. The flow downstream of the separation shock-wave is apparently uniform and the edge of the boundary layer is therefore nearly straight. The dark line in the white boundary layer downstream of the separation point is due to 'overloading' of the schlieren system (see Mair (1952)) and could be made to disappear by increasing the width of the second slit. It should be noted, however, that it may be advantageous to adjust the schlieren system so that the dark line occurs, as it sometimes makes it easier to interpret the photographs. The overloading indicates a strong density gradient in a layer separated from the surface. In fig. 10 the vortex sheet starting at the point of intersection of the two strong shock waves can be seen.

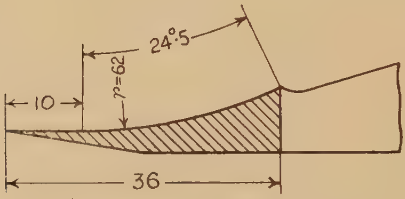
---

* For plates see end of issue.

By measuring the photographs, it was found that the slope of the first shock wave agreed with inviscid theory within $\frac{1}{2}^\circ$. The second shock wave, where the boundary layer separated, gave a deflection of approximately 5°. The position of the point of separation was 6 to 7 mm upstream of the corner. This is approximately 60 times the displacement thickness of a boundary layer on a flat plate at the Mach number and Reynolds number of the experiments and at the same distance from the leading edge as the corner. The angle of the main shock (1–5) agreed within 0·2–0·5° with the value calculated for a total deflection of $(\delta_1+\delta_2)$.

The surface was then made rough by spraying it with 'dead black' paint, and the photographs figs. 11 and 12 (Plate XXV) were obtained for $\delta_1=0{\cdot}0^\circ$ and $\delta_1=7{\cdot}5^\circ$ respectively, with δ_2 equal to 9·4° in both cases. In fig. 12 the upstream effect of the corner shock-wave is hardly detectable. In fig. 11 there is still some upstream effect, but much less than in the case of the smooth surface. It is believed that the difference in magnitude of the upstream effects shown in figs. 11 and 12 is not due to the difference in Mach number upstream of the corner shock-wave, but is caused by the greater roughness of the surface in fig. 12. The method used for roughening the surface had the disadvantage that the paint was blown off in the course of a few seconds, and it was therefore not possible to get exactly the same roughness in the two cases. In both figures the wavelets from the rough surface are clearly visible. The irregular wavy line below the model in fig. 12 is a water smear running downstream along the glass window, and it therefore represents a streamline in the boundary layer on the window. Note that the trace bends before reaching the image of the shock wave from the leading edge, because the shock wave bends upstream when it runs into the boundary layer. The water smear was caused by a drop of dirty water issuing from a cavity in the upstream joint of the glass window. The water drop remained there after the window had been washed and dried, but it could only have a momentary effect on the dryness of the air in the main stream.

Fig. 13



Section of circular-arc plate (with dimensions in mm).

The circular-arc profile had the dimensions shown in fig. 13. A flat part upstream of the curved part prevented the shock wave due to the slight bluntness of the leading edge (see Bardsley (1951)) from obscuring the flow over the curved profile. The total deflection of a streamline along the surface of the profile was 24·5°. With a smooth surface the flow was photographed with the straight part set at incidences 0·2°, 2·7° and 5·1°. These incidences correspond to Mach numbers just upstream of the leading Mach line of the curved surface (i.e. in the effective free stream) of 1·97, 2·05 and 2·15.

Figures 14 and 15 (Plate XXVI) show the flow for $M_1=1{\cdot}97$ and fig. 16 (Plate XXVI) the flow for $M_1=2{\cdot}15$. As on the double wedge the compression is propagated upstream, the boundary layer separates, and a shock wave occurs at the point of separation. Again the flow downstream of the separation shock-wave is apparently uniform, and the edge of the boundary layer is therefore nearly straight.

By measuring the photographs, it was found that the shock-wave angles agreed well with the values calculated from the measured slopes of the boundary layer. The total deflection of the streamlines (outside the boundary layer) was found to be 22·5°, i.e. 2° less than that of a streamline at the surface. This is because the expansion at the downstream edge of the model produces a negative pressure gradient in the boundary layer on the upstream side, so reducing the rate of growth of the boundary layer. For the decaying shock wave it was found that the distance between the computed wave and the one found experimentally* was only about 1 mm on the scale of the experiments. This corresponds to about 1/50 of the radius of the circular arc.

The deflection at the point of separation was approximately 4°, and there was a further deflection of about 4° at the point of reattachment. The point of separation was about 2 mm upstream of the beginning of the circular arc. This is about 20 times the displacement thickness of an undisturbed boundary layer at the beginning of the curved part.

When the surface was made rough the photograph fig. 17 (Plate XXVI) was obtained for $M_1=1{\cdot}96$ (zero incidence). The upstream effect is no longer detectable and the flow everywhere outside the boundary layer is in good agreement with inviscid theory (fig. 6).

In all the photographs figs. 14–17 the flow corresponds *roughly* to the case of a centred compression meeting a single strong shock wave, and the weak shock wave in the direction of the downstream Mach line can be seen in figs. 14 and 15. It could also be seen on the negatives from which figs. 16 and 17 were reproduced.

§5. CONCLUSIONS

The supersonic flow over a double wedge (fig. 8) and a circular-arc profile (fig. 13) was photographed and the actual flow compared with that predicted by inviscid theory.

It was found that when the surface of the profile was rough and the boundary layer therefore turbulent (e.g. see figs. 12 and 17), the flow outside the boundary layer was in good agreement with the inviscid flow.

---

* The downstream side of the image represents the true position of the shock wave outside the boundary layer (see Bardsley and Mair (1951)).

When the surface was smooth and the boundary layer laminar the increase of pressure on the profile was propagated upstream through the boundary layer and caused separation (figs. 10 and 14). A shock wave originated at the point of separation and caused considerable deviation from the inviscid flow. The flow at points far from the surface was in all cases in good agreement with the inviscid flow, whatever the state of the boundary layer.

ACKNOWLEDGMENTS

The author thanks Professor M. J. Lighthill and Mr. W. A. Mair for their valuable suggestions, and for the interest they have taken in the work.

Part of the work was carried out while the author was assisted by grants from the British Council and from K. A. Larssens og Hustru L. M. Larssen født Thodbergs Legat.

REFERENCES

BARDSLEY, O., 1951, *Phil. Mag.*, **42,** 255–262.
BARDSLEY, O., and MAIR, W. A., 1951, *Phil. Mag.*, **42,** 29–36.
COURANT, R., and FRIEDRICHS, K. O., 1948, *Supersonic Flow and Shock Waves, Interscience*, p. 464.
FERRI, A., 1949, *Elements of Aerodynamics of Supersonic Flows.* Macmillan. 434 pp.
HERBERT, P. J., and OLDER, S. J., 1946, *A.R.C.*, 10261.
KAHANE, A., and LEES, L., 1948, *J. Aeron. Sci.*, **15,** 167–170.
MAIR, W. A., 1952, to be published in the *Aeron. Quart.*
MEYER, R. E., 1948 a, *Quart. J. Mech. and Appl. Math.*, **1,** 196–219 ; 1948 b, *Ibid.*, 451–469.

LIII. *The Spatial Correlation of Electrons in Atoms and Molecules: I Helium and Similar Two Electron Systems in their Ground States*

By Sir John Lennard-Jones, F.R.S., and J. A. Pople
Department of Theoretical Chemistry, University of Cambridge*

[Received February 24, 1952]

Summary

An attempt is made in this paper to improve the theory of two electrons of opposite spin in a field of spherical symmetry such as exists in some atoms. The method follows closely the theory of paired electrons in a chemical bond, given recently by the authors. In order to take fully into account the spatial correlation of the electrons due to their electrostatic repulsion, the wave function is expressed as a series. Each member of the series contains a function of the radial distances of the electrons multiplied by a Legendre function of the angle which the electrons subtend at the nucleus. It is then found possible to obtain a series of coupled equations for the radial functions in which the angular coordinate does not appear. The exact theory is used to examine some approximate expressions which have been used for the wave functions of the helium atom.

Spatial correlation may occur in two ways. It may be of the 'in-out' type, one electron tending to be outside the other, or it may be of the angular type, the two electrons tending to be on opposite sides of the nucleus. The spatial correlation of the electrons in helium in its ground state is examined from this point of view.

§1. Introduction

One of the greatest difficulties in the theory of atomic and molecular structure is to take adequately into account the electrostatic repulsion of the electrons. Because of this interaction it is not possible to describe satisfactorily the properties of many electron systems in ordinary three-dimensional space. Even a two-electron system such as a helium atom in its normal state has to be described in six dimensions (or eight dimensions, if the spin properties are included). This is because there is a correlation between the behaviour of the electrons and in order to describe the system fully, it is necessary to know the probability of finding simultaneously one electron in one element of space and a second electron in another prescribed element of space. It has not as yet been possible to devise an experiment which would disclose this inner property of an atom. All that diffraction experiments, for example, can demonstrate is the average total distribution in three dimensions. This total electron density of a helium atom is known to be spherically symmetrical. The hands of a clock provide a simple analogy. Each hand has on the average the same probability of being in a prescribed element of angle, and if a photograph were taken with a long exposure the image of the two hands would have circular symmetry. But there is another quantity which could be deduced about the hands, and that is the probability of finding them at any prescribed angle to each other. So in an atom or a molecule there is a probability of finding the electrons in given configurations relative to each other.

---

* Communicated by the Authors.

The correlation of electrons in molecules is of more direct interest than in atoms because it is a major factor in determining molecular shape. It is for this reason a subject of great theoretical interest. It is accordingly instructive to examine the spatial correlation of electrons in atoms if only as a guide to the more complex systems which exist in molecules.

While the spatial correlation of electrons in atoms is not itself directly observable, it none the less affects other observable properties. The energy content of atoms depends directly on the average distance apart of the electrons. Attempts to calculate atomic energies involve implicitly, if not explicitly, methods of estimating, or of guessing, the distribution of electrons relative to each other. Many attempts have been made to calculate the energy of two electron systems, like helium, by variational methods, but the primary emphasis has been on such observables as the ionization potential and not on the distribution of the electrons relative to each other.

In this paper the effect of the electrostatic repulsion on the wave function of a pair of electrons in an atom, when this has a spherically symmetrical field, has been examined by a method similar to that used elsewhere for a pair of electrons in a chemical bond (Lennard-Jones and Pople 1951). The wave function is expanded in a series similar to that used for the electrostatic repulsion when expressed in terms of the radial coordinates of the two electrons and the angle they subtend at the nucleus. The dependence on this angle then appears in terms of spherical harmonics, and it is necessary only to find how the wave function depends on the radial coordinates. It is shown how in principle an exact solution of this problem can be obtained. This theory is then used to examine the limitations of various approximate methods which have been used, and in particular to show the significance of the calculations of Taylor and Parr of the energy of the helium atom (Taylor and Parr 1952).

For paired electrons in the lowest energy state of a system like helium it is the electrostatic repulsion which is responsible for spatial correlation. This is always operative whether electrons have the same spin or opposite spin. When electrons have the same spin, there is another factor which may be more important, and that is the exclusion principle. Though not of direct concern in this paper, the great influence of the exclusion principle on the spatial correlation of electrons must for the sake of completeness be pointed out. This principle is usually expressed mathematically in the form that the total wave function of a many-electron system must be antisymmetric for interchanges of electrons. The significance of the principle is most easily appreciated by considering the fictitious system of many electrons moving without mutual repulsion in the same field, whether atomic or molecular. The wave function of this system can then be expressed accurately as a product, or sum of products, of one electron functions. For the lowest state the wave function has the form of a determinant. The exclusion principle then demands that not more than one electron of the same spin be assigned to the same one electron function. The wave function vanishes when any two electrons have the same spin and the same spatial coordinates. This means that electrons of the same spin avoid each other and tend to occupy different regions of space. The significance of this has been examined for a number of atoms and molecules in another paper (Lennard Jones 1952).

§2. THE SEPARATION OF THE WAVE EQUATION

The Hamiltonian of a system of two electrons moving in a spherically symmetric field V is, in atomic units,

$$\mathfrak{H}=H_1+H_2+1/r_{12}, \tag{2.1}$$

where

$$\left.\begin{aligned} H_1&=-\tfrac{1}{2}\nabla_1^2+V(r_1),\\ H_2&=-\tfrac{1}{2}\nabla_2^2+V(r_2).\end{aligned}\right\} \tag{2.2}$$

In these formulae r_1 and r_2 are the distances of the two electrons from the centre of the symmetrical field and r_{12} is the interelectronic distance. V may represent the field due to a central nucleus, as in the helium atom or the Li^+ ion, or it may represent the combined effect of a nucleus and the screening by a closed shell of electrons, as in atoms such as beryllium and magnesium.

The interelectronic potential $(1/r_{12})$ depends only on r_1, r_2 and θ_{12}, the angle subtended at the centre by the two electrons. It may be expanded in terms of Legendre functions in the form

$$1/r_{12}=\sum_{l=0}^{\infty}(r_<^l/r_>^{l+1})P_l(\cos\theta_{12}), \tag{2.3}$$

where $r_<$ and $r_>$ are the smaller and greater of r_1 and r_2 respectively, and where P_l are Legendre polynomials defined so that $P_l(1)=1$. For brevity this will be written

$$1/r_{12}=\sum_{l=0}^{\infty}U_l(r_1,r_2)P_l(\cos\theta_{12}), \tag{2.4}$$

U_l being independent of θ_{12}.

The solutions of the wave equation, when the Hamiltonian has the form (2.1), will be of various symmetry types. For those of symmetry ^1S it is convenient to expand ψ in a series similar to (2.4). Thus we write

$$\psi(1,2)=\sum_{l=0}^{\infty}\psi_l(r_1,r_2)P_l(\cos\theta_{12}), \quad . \quad . \quad . \quad . \quad . \quad (2.5)$$

ψ_l being a function of r_1 and r_2 only.

The expansions (2.4) and (2.5) are independent of the choice of a fixed reference frame for specifying angular position, but they can easily be transformed to expressions involving surface harmonics referred to such a frame. If the spherical polar coordinates of the two electrons are (r_1, θ_1, ϕ_1) and (r_2, θ_2, ϕ_2) respectively, then*

$$P_l(\cos\theta_{12})=P_l(\cos\theta_1)P_l(\cos\theta_2)$$
$$+2\sum_{m=1}^{l}\frac{(l-m)!}{(l+m)!}P_l^m(\cos\theta_1)P_l^m(\cos\theta_2)\cos m(\phi_2-\phi_1). \quad (2.6)$$

Using (2.5) the last term in the Schrödinger equation becomes

$$(1/r_{12})\psi=\sum_{l,m}^{\infty}U_l\psi_m P_l(\cos\theta_{12})P_m(\cos\theta_{12}). \quad . \quad . \quad . \quad (2.7)$$

A product of Legendre polynomials can be expanded as a sum by the formula (Adams 1878)

$$P_l(\mu)P_m(\mu)=\sum_{r=0}^{\mathrm{Min}(l,m)}c_{lm}{}^{(r)}P_{l+m-2r}(\mu), \quad . \quad . \quad . \quad . \quad (2.8)$$

where the $c_{lm}{}^{(r)}$ are numerical coefficients given by

$$c_{lm}{}^{(r)}=\frac{A_{l-r}A_rA_{m-r}}{A_{l+m-r}}\,\frac{2l+2m-4r+1}{2l+2m-2r+1}, \quad . \quad . \quad . \quad . \quad (2.9)$$

and

$$A_s=\frac{1\,.\,3\,.\,5\ldots(2s-1)}{s\,!}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2.10)$$

Using this formula, the complete Schrödinger equation can now be separated into a set of coupled equations for the functions ψ_l, since the coefficients of the various Legendre functions must separately vanish. Thus we get

$$\left\{-\tfrac{1}{2}\left[\frac{1}{r_1^2}\frac{\partial}{\partial r_1}\left(r_1^2\frac{\partial}{\partial r_1}\right)+\frac{1}{r_2^2}\frac{\partial}{\partial r_2}\left(r_2^2\frac{\partial}{\partial r_2}\right)\right]+\frac{l(l+1)}{2}\left[\frac{1}{r_1^2}+\frac{1}{r_2^2}\right]\right.$$
$$\left.+V(r_1)+V(r_2)\right\}\psi_l+\sum_{mn}c_{mn}^{\frac{1}{2}(m+n-l)}U_m\psi_n=E\psi_l, \quad . \quad . \quad (2.11)$$

where the $c_{mn}{}^{(r)}$ are defined to be zero if r is negative or non-integral.

---

* For states of the system which do not possess spherical symmetry it will be necessary to use (2.4) together with (2.6) and to expand the wave function in a form similar to (2.6), each term in the expansion being multiplied by a function of r_1 and r_2. It is hoped to consider this problem in a later paper.

We note, in particular, that the equation for ψ_0 takes the form

$$\left\{-\tfrac{1}{2}\left[\frac{1}{r_1{}^2}\frac{\partial}{\partial r_1}\left(r_1{}^2\frac{\partial}{\partial r_1}\right)+\frac{1}{r_2{}^2}\frac{\partial}{\partial r_2}\left(r_2{}^2\frac{\partial}{\partial r_2}\right)\right]+V(r_1)+V(r_2)\right\}\psi_0 + \sum_{n=0}^{\infty}\frac{U_n}{2n+1}\psi_n=E\psi_0. \qquad (2.12)$$

Thus the purely radial function ψ_0 depends on the coefficients of the higher angular terms in the full expression for $\psi(1, 2)$.

Up to this point the treatment is exact except for any approximation implicit in the use of a smoothed potential function V. The eqns. (2.11) for ψ_l may not be soluble explicitly, but as in the case of a pair of electrons in a chemical bond (Lennard-Jones and Pople 1951), the accurate theory forms a useful background against which to examine the limitations of approximate wave functions.

§3. Approximate Solutions in Terms of One-electron Functions

In order to examine the physical significance of approximate wave functions, we shall consider first those which are independent of the angle θ_{12}. This implies neglecting all but the first term ψ_0 in the expansion (2.5). The simplest approximation to ψ_0 with the necessary symmetry properties is a product of one-electron functions,

$$\psi=f_0(r_1)f_0(r_2). \qquad (3.1)$$

If f_0 is chosen so as to minimize the energy integral $\int\bar{\psi}\mathfrak{H}\psi\,d\tau\,/\int\bar{\psi}\psi\,d\tau$, then this is the simple Hartree-Fock function. According to this function each electron moves in the averaged field of the other, but otherwise there is no correlation between them. No allowance is made for their tendency to avoid one another because of their electrostatic repulsion.

The function ψ_0 cannot, of course, allow for angular correlation, that is the tendency for the angle θ_{12} to be as large as possible. More accurate forms than (3.1) do, however, introduce radial correlation, the tendency for the two electrons to be at different radii. As a better approximation we may accordingly replace (3.1) by

$$\psi=f_{0,a}(r_1)f_{0,a}(r_2)-f_{0,b}(r_1)f_{0,b}(r_2). \qquad (3.2)$$

The negative sign is used in (3.2) as it is found later that this avoids the use of imaginary functions. As we are only interested in the function ψ_0, we shall drop the zero suffix for the rest of this section. We then have

$$\int\bar{\psi}\psi\,d\tau=S_{aa}{}^2+S_{bb}{}^2-2S_{ab}{}^2, \qquad (3.3)$$

and

$$\int\bar{\psi}\mathfrak{H}\psi\,d\tau=2(S_{aa}H_{aa}+S_{bb}H_{bb}-2S_{ab}H_{ab}) +\{(aa|U_0|aa)+(bb|U_0|bb)-2(aa|U_0|bb)\}, \qquad (3.4)$$

where we have written

$$\left.\begin{aligned} S_{aa}&=\int\bar{f}_af_a\,dx,\\ H_{aa}&=\int\bar{f}_aHf_a\,dx,\\ (aa|U_0|bb)&=\iint\bar{f}_a(1)\bar{f}_a(2)U_0f_b(1)f_b(2)\,dx_1\,dx_2, \end{aligned}\right\} \qquad (3.5)$$

and other similar expressions.

To obtain the optimum forms for f_a and f_b we have to minimize the integral $\int \bar{\psi}\mathfrak{H}\psi\, d\tau$, subject to the condition

$$\int \bar{\psi}\psi\, d\tau = 1. \qquad (3.6)$$

This leads to the following pair of coupled integro-differential equations

$$\left.\begin{aligned} H\{S_{aa}f_a - S_{ab}f_b\} + \{(U_0)_{aa}f_a - (U_0)_{ab}f_b\} \\ = (ES_{aa} - H_{aa})f_a + (ES_{ab} - H_{ab})f_b, \\ H\{S_{bb}f_b - S_{ab}f_a\} + \{(U_0)_{bb}f_b - (U_0)_{ab}f_a\} \\ = (ES_{bb} - H_{bb})f_b + (ES_{ab} - H_{ab})f_a, \end{aligned}\right\}, \qquad (3.7)$$

where we have written

$$(U_0)_{aa} = \int \bar{f}_a(2) U_0 f_a(2)\, dx_2, \qquad (3.8)$$

and other similar expressions. The solution of (3.7) must be scaled so as to satisfy the normalization condition

$$S_{aa}^2 + S_{bb}^2 - 2S_{ab}^2 = 1. \qquad (3.9)$$

We note that if we put $f_b = 0$ in the first of eqns. (3.7), the ordinary Hartree-Fock equation is obtained. There are two distinct approximations, therefore, in this function, one being the expression of ψ_0 as a simple product in the form (3.1) and the other the neglect of all the terms in (2.12) other than ψ_0.

An alternative form of the function (3.2) can be obtained by transforming to new one-electron orbitals given by

$$\left.\begin{aligned} \chi &= (f_a + f_b)/\sqrt{2}, \\ \chi' &= (f_a - f_b)/\sqrt{2}, \end{aligned}\right\} \qquad (3.10)$$

so that the wave function becomes

$$\psi = \chi(1)\chi'(2) + \chi'(1)\chi(2). \qquad (3.11)$$

To obtain further improvement it is necessary to include approximations to the functions ψ_l for $l \geqslant 1$. Such approximations may be simple products as in (3.1), giving a total wave function

$$\psi = \sum_{l=0}^{\infty} f_l(r_1) f_l(r_2) P_l(\cos\theta_{12}), \qquad (3.12)$$

or sums of two simple products as in (3.2)

$$\psi = \sum_{l=0}^{\infty} \{f_{l,a}(r_1) f_{l,a}(r_2) - f_{l,b}(r_1) f_{l,b}(r_2)\} P_l(\cos\theta_{12}). \qquad (3.13)$$

Equations for the optimum forms of any of these functions can be obtained by minimizing the integral $\int \bar{\psi}\mathfrak{H}\psi\, d\tau$ subject to the normalization condition, using the corresponding form for ψ. It is to be noted that the wave function (3.13) can be written as a generalization of (3.11) in the form

$$\psi = \sum_{l=0}^{\infty} \{\chi_l(1)\chi_l'(2) + \chi_l'(1)\chi_l(2)\} P_l(\cos\theta_{12}). \qquad (3.14)$$

§4. ILLUSTRATIVE NUMERICAL RESULTS FOR THE HELIUM ATOM

In order to illustrate the way in which the introduction of the higher components of the wave function lowers the total calculated energy by allowing for the tendency of the electrons to avoid one another, it is instructive to examine some results obtained recently for the helium atom by Russell Taylor and Parr (1952).* Their calculations are based on hydrogen-like atomic functions instead of on the one-electron orbitals discussed above. By the use of such functions the calculations can be done simply, but it should be emphasized that there is no reason to suppose that they are good approximations to the optimum forms. The functions considered are the following :—

Function 1. An approximation to $\psi_0(r_1, r_2)$ of the form $f_0(1)f_0(2)$ using a hydrogen-like $(1s)$ function (with the best value of Z) for f_0. The optimum value of Z is 1.6875.

Function 2. An approximation to $\psi_0(r_1, r_2)$ of the form $\{\chi_0(1)\chi_0'(2)+\chi_0'(1)\chi_0(2)\}$ using hydrogen-like $(1s)$ functions, with different screening constants (Z and Z') for χ_0 and χ_0'. The values of Z and Z' are 1.19 and 2.18365.

Function 3. To function 2 is added an expression $f_1(1)f_1(2)P_1(\cos\theta_{12})$ using a function proportional to the radial part of a hydrogen-like $2p$-function for f_1. Expanding $P_1(\cos\theta_{12})$ by the addition theorem for Legendre polynomials, it is seen that function 3 corresponds to the superposition of the configurations $(1s, 1s')$ and $(2p)^2$. The screening constant in the radial function f_1, is taken to be 4.9504.

Function 4. Similar approximate forms are added to represent the effect of $(3d)$ and $(4f)$ functions. The full expression used can be written in the form

$$\begin{aligned}\psi=\{\chi_0(1)\chi_0'(2)+\chi_0'(1)\chi_0(2)\}&+f_1(1)f_1(2)P_1(\cos\theta_{12})\\&+f_2(1)f_2(2)P_2(\cos\theta_{12})+f_3(1)f_3(2)P_3(\cos\theta_{12}). \quad . \quad . \quad (4.1)\end{aligned}$$

The functions f_2 and f_3 are taken to be proportional to the radial parts of hydrogen-like $(3d)$ and $(4f)$ functions. The screening constants in the radial functions f_2 and f_3 are taken to be 11.305 and 19.04.

The energies of the helium atom calculated from these functions are shown in table 1, which is taken from the paper of Taylor and Parr (1952).

Table 1. Calculated Energies for the Normal Helium Atom

	$-E_{\text{calc}}$ (units of $e^2/2a_0$)	$E_{\text{calc}}-E_{\text{expt}}$ (units of $e^2/2a_0$)
Function 1.	5·69531	0·1113
Function 2.	5.751316	0·0552
Function 3.	5·790369	0·0162
Function 4.	5·794860	0·0117

---

* We are indebted to the authors for allowing us to refer to their results before publication.

As Taylor and Parr pointed out, these figures illustrate the relative improvements obtained by allowing for the various types of correlation. The analysis of this paper shows the relation of these improvements to the properties of the exact wave function (2.5).

§5. The Spatial Correlation of Electrons in the Helium Atom

Some further insight into the correlation between the electrons in the helium atom can be obtained by a more detailed examination of the various parts of the wave function. It should be emphasized, however, that any conclusions based on the approximate functions of Taylor and Parr may be inaccurate because of the particular types of radial functions they have used.

In terms of the accurate function (2.5), the probability of the radial distance of the first electron being between r_1 and r_1+dr_1, of the radial distance of the second electron being between r_2 and r_2+dr_2 and of the angle subtended at the nucleus being between θ_{12} and $\theta_{12}+d\theta_{12}$ is

$$P(r_1, r_2, \theta_{12})\, dr_1\, dr_2\, d\theta_{12},$$

where

$$P(r_1, r_2, \theta_{12}) = \left[\sum_{l=0}^{\infty} \psi_l^2 \{P_l(\cos\theta_{12})\}^2 + 2 \sum_{m>n}^{\infty} \psi_m \psi_n P_m(\cos\theta_{12}) P_n(\cos\theta_{12}) \right] 8\pi^2 r_1^2 r_2^2 \sin\theta_{12}. \qquad (5.1)$$

To investigate *radial* correlation irrespective of *angular* correlation in the helium atom, we must integrate (5.1) over all values of θ_{12} between 0 and π. This leads to a radial probability distribution function $P(r_1, r_2)\, dr_1\, dr_2$ where

$$P(r_1, r_2) = 16\pi^2 r_1^2 r_2^2 \sum_{l=0}^{\infty} \frac{1}{2l+1} \psi_l^2. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5.2)$$

From the form of (5.2) it is clear that the radial correlation will be mainly described by the first term of the wave function ψ_0, partly because this is the largest term in the expansion (2.5) and partly because of the numerical factors $(2l+1)^{-1}$.

If we neglect all but the first term in (5.2) and then use the particular approximation to ψ_0 given by function 2 of the last paragraph where

$$\left.\begin{aligned} \chi_0 &\propto e^{-Zr}, \\ \chi_0' &\propto e^{-Z'r}, \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5.3)$$

then $P(r_1, r_2)$ is proportional to the function

$$F(s, t) = 4(s^2 - t^2)^2\, e^{-2(Z+Z')s} [\cosh\{(Z'-Z)t\}]^2, \quad . \quad . \qquad (5.4)$$

where

$$\left.\begin{aligned} s &= \tfrac{1}{2}(r_1 + r_2), \\ t &= \tfrac{1}{2}(r_1 - r_2). \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (5.5)$$

It is of interest to find out under what circumstances the values of r_1 and r_2 tend to differ. From (5.4) we obtain

$$\left[\frac{\partial^2 F}{\partial t^2}\right]_{t=0} = 8s^2\{(Z'-Z)^2s^2-2\}\,\mathrm{e}^{-2(Z+Z')s} \qquad (5.6)$$

and this gives the sign of the curvature of the surface F along the section $r_1=r_2$. If s is large, this will be positive. In other words, if the *average* of the two radii is large, the two electrons will tend to separate into different spherical shells and radial correlation of the 'in-out' type will occur. But if s is small, (5.6) will be negative and the electrons will most probably be found in the same spherical shell. The critical value separating the two cases is

$$s_{\text{crit}} = \sqrt{2}/|Z'-Z|. \qquad (5.7)$$

Using the values of Z' and Z given by Taylor and Parr, this corresponds to about 0.75 Å.

It is more illuminating to inquire what is the most probable value of r_2 when r_1 is given. For function 1 of the last section this quantity is independent of r_1, there being then no radial correlation. It appears that for function 2, given by (5.3), the most probable value of r_2 is greater or less than a certain value r_0 according as r_1 is less or greater than this quantity. For the values of Z and Z' given above for function 2 the value of r_0 proves to be about 0.3 Å. The most probable value of r_2 ranges from about 0.34 Å to 0.23 Å as r_1 changes from small to large values.

To examine angular correlation, we must return to the complete distribution function (5.1). The importance of angular correlation for given values of r_1 and r_2 may be assessed quantitatively in several ways. One way would be to compare the probabilities per unit solid angle of the two configurations $(r_1, r_2, 0)$ and (r_1, r_2, π), that is the configurations in which the two electrons lie on the same and opposite sides of the nucleus along a given radius. The ratio of these two probabilities is given by

$$\frac{P(r_1, r_2, 0)}{P(r_1, r_2, \pi)} = \frac{\sum\limits_{l=0}^{\infty} \psi_l^2 + 2\sum\limits_{m>n}^{\infty} \psi_m\psi_n}{\sum\limits_{l=0}^{\infty} \psi_l^2 + 2\sum\limits_{m>n}^{\infty} (-1)^{m+n}\psi_m\psi_n}. \qquad (5.8)$$

If we neglect all functions ψ_l for $l>1$, and then use the approximation to $\psi_0+\psi_1\cos\theta_{12}$ given by function 3 of paragraph 4, with

$$f_1 \propto r\mathrm{e}^{-Z''r/2}, \qquad (5.9)$$

then (5.8) becomes

$$\frac{P(r_1, r_2, 0)}{P(r_1, r_2, \pi)} = \left\{\frac{1+\psi_1/\psi_0}{1-\psi_1/\psi_0}\right\}^2, \qquad (5.10)$$

where

$$\begin{aligned}\psi_1/\psi_0 &= \kappa\,\frac{r_1 r_2\,\mathrm{e}^{-Z''(r_1+r_2)/2}}{\mathrm{e}^{-Zr_1-Z'r_2}+\mathrm{e}^{-Zr_2-Z'r_1}} \\ &= \tfrac{1}{2}\kappa(s^2-t^2)\,\mathrm{e}^{-(Z''-Z-Z')s}\,\operatorname{sech}\{(Z'-Z)t\}, \qquad (5.11)\end{aligned}$$

κ being a (negative) constant. For a given value of s, the expression (5.11) is always numerically largest when $t=0$, that is when $r_1=r_2$. In other words, angular correlation is of greatest importance when the two electrons are at the same radii, a result which is not surprising.

If we put $t=0$, we find that (5.11) has a maximum when

$$r_1=r_2=2/(Z''-Z-Z'). \qquad (5.12)$$

This is the radius at which angular correlation is most important. Using the results of Taylor and Parr for the Z's, this is about 0.67 Å.

Alternatively, we can estimate the importance of angular correlation by calculating the mean value of $\cos\theta_{12}$ given the two radial distances r_1 and r_2. This is given by

$$\overline{(\cos\theta_{12})}_{r_1,r_2}=\frac{\int_0^\pi \cos\theta_{12}P(r_1,r_2,\theta_{12})\,d\theta_{12}}{\int_0^\pi P(r_1,r_2,\theta_{12})\,d\theta_{12}}$$

$$=\sum_{l=0}^{\infty}\frac{2(l+1)}{(2l+1)(2l+3)}\psi_l\psi_{l+1}\Big/\sum_{l=0}^{\infty}\frac{1}{(2l+1)}\psi_l^2. \qquad (5.13)$$

If we omit all ψ_l with $l>1$, we get

$$\overline{(\cos\theta_{12})}_{r_1,r_2}=\tfrac{2}{3}\psi_0\psi_1/(\psi_0^2+\tfrac{1}{3}\psi_1^2). \qquad (5.14)$$

To a good approximation, the denominator can be replaced by ψ_0^2 and (5.15) then reduces to $\frac{2}{3}\psi_1/\psi_0$, and so is proportional to (5.11). It is of interest to calculate the numerical value of this quantity for two special cases. In the absence of spatial correlation it would, of course, be zero and the average value of θ_{12} would be 90°.

(i) If $r_1=r_2=2/(Z''-Z-Z')$, corresponding to both electrons being at the radius (0.67 Å) at which angular correlation is most important, then $\overline{(\cos\theta_{12})}_{r_1,r_2}=-0.1073$ (or $\theta_{12}=96^\circ$ approximately).

(ii) If $r_1=r_2=2/(Z+Z')$ corresponding to both electrons being at the most probable radius (about 0.3 Å), then $\overline{(\cos\theta_{12})}_{r_1,r_2}=-0.0687$ (or $\theta_{12}=94^\circ$ approximately).

Finally, we may calculate the mean value of $\cos\theta_{12}$ averaged over the whole of configuration space. This is given by

$$\overline{\cos\theta_{12}}=\int_0^\infty\int_0^\infty\int_0^\pi \cos\theta_{12}P(r_1,r_2,\theta_{12})\,dr_1\,dr_2\,d\theta_{12}$$

$$=\sum_{l=0}^{\infty}\frac{32\pi^2(l+1)}{(2l+1)(2l+3)}\int_0^\infty\int_0^\infty r_1^2r_2^2\psi_l\psi_{l+1}\,dr_1\,dr_2. \qquad (5.15)$$

With the same approximations as before, this is estimated to be -0.0665 corresponding to an angle of 93.8°.*

---

* We are indebted to Mr. Bhattacharya and to Mr. A. Brickstock for some of the numerical values given in this section.

§6. APPLICATION TO ATOMIC AND MOLECULAR STRUCTURE

In this paper we have been primarily concerned with the spatial correlation of two electrons resulting from their electrostatic repulsion. It has already been noted in the introduction, however, that for systems with more than two electrons, it is also necessary to take into account the spatial correlation that arises from the exclusion principle. The effect of the latter has already been considered for a number of atoms and molecules (Lennard-Jones 1952).

The next step in the theory must clearly be a comparison of the relative importance of the two causes of correlation in a system in which both operate. Thus, for the neon atom, the exclusion principle tends, to a first approximation, to keep the four outer electrons of the same spin at the four corners of a regular tetrahedron (Lennard-Jones 1952). If the exclusion principle were the only factor influencing the spatial correlation, the two tetrahedra so defined would move independently of one another. But if the correlation due to electrostatic repulsion is allowed for, there will be a tendency for the two tetrahedra to keep apart, and it seems likely that the most probable configuration of the eight outer electrons will be that in which each is situated at the corner of a cube. The calculations of this paper do indicate, however, that the effect of electrostatic repulsion on angular correlation is comparatively small, so the relative motion of the two tetrahedra may be much freer than that of the electrons within the tetrahedron. It is hoped to publish further work on this problem later.

There are other important applications of these ideas in the field of molecular structure. In the paper dealing with the effect of the exclusion principle (Lennard-Jones 1952), it was shown that three electrons of the same spin in a triple bond, as in the nitrogen and acetylene molecules, are most probably in azimuthal planes inclined to each other at angles of 120°. If we ignore spatial correlation due to electrostatic repulsion (as we do if we use a single determinant wave function), the two systems of planes thus defined rotate about the axis of the bond independently. Only further development of calculations of the type discussed in this paper can lead to further knowledge of the extent to which the six electrons are kept in azimuthal planes inclined to each other at angles of 60°.

REFERENCES

ADAMS, J. C., 1878, *Proc. Roy. Soc.*, **27**, 63.

LENNARD-JONES, Sir J., 1952, *J. Chem. Phys.* (in the press).

LENNARD-JONES, Sir J., and POPLE, J. A., 1950, *Proc. Roy. Soc.* A, **202**, 166; 1951, *Ibid.*, **210**, 190.

TAYLOR, G. R., and PARR, R. G., 1952, *Proc. Nat. Acad. Sci.* (in the press).

LIV. CORRESPONDENCE

A New Nuclear-Research Emulsion

By A. J. Herz
Imperial College, London, S.W.7
and
C. Waller
Ilford Ltd., Ilford, Essex*

[Received March 11, 1952]

Nuclear-Research emulsions are often used to investigate the properties of the heavy nuclei in the cosmic radiation. Several ways of identifying these particles have been proposed and used ; they involve a determination of the rate of energy loss combined with that of the residual range or scattering. If a particle is known to be moving with relativistic velocity ($v/c \approx 1$), its rate of energy loss is proportional to the square of its charge, and identification is then possible by the measurement of energy loss alone.

In most investigations, emulsion sensitive to all ionizing particles is needed ; as a result the tracks of heavy nuclei are continuous and it is impossible to determine the rate of energy loss directly by grain counting. This difficulty has been overcome in two ways. Firstly, it is possible to measure the density of delta-rays along the track, a quantity which is closely related to the total ionization (Bradt and Peters 1948, Hoang 1950, Sørensen 1949), or the widening of the core of the track produced by low-energy delta-rays (Blau, Rudin and Lindenbaum 1950, Ceccarelli and Zorn 1952). The second method is to interleave the sensitive plates with very insensitive ones (Bradt and Peters 1950). Grain counting is then possible on the tracks of heavy nuclei, and the rate of energy loss can be found directly. The advantages of this method are that grain counting is simple and quick, that the statistical accuracy for a given length of track is greater than that of delta-ray counts, and that grain counting is not as subjective a measurement as delta-ray counting so that the establishment and maintenance of elaborate conventions is unnecessary. The disadvantage of grain counting is that every batch of plates has to be calibrated separately. Other disadvantages of the method of grain counting as used by Bradt and Peters (1950) are that the insensitive plates have to be developed separately by special methods in order to obtain the required low sensitivity, and that the tracks have to be traced through a stack of plates.

---

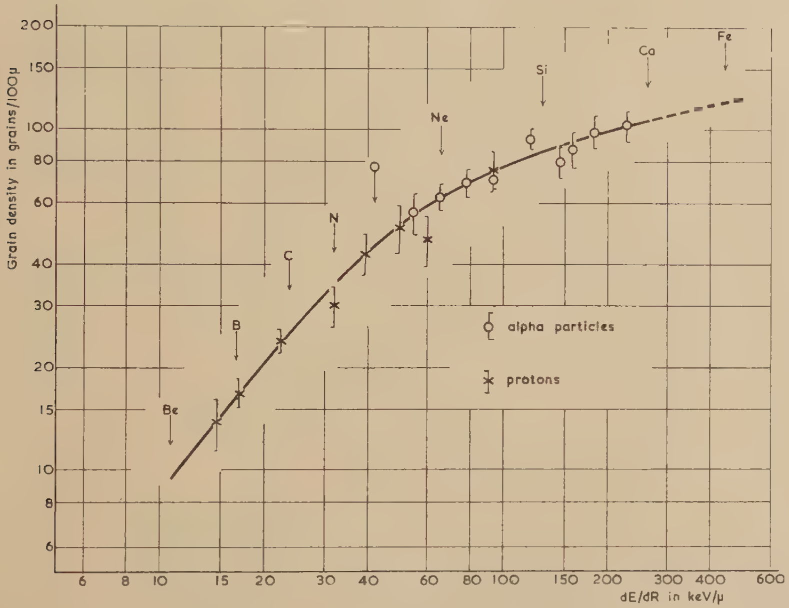
*Communicated by Sir George Thomson, F.R.S.

To overcome the disadvantages of the grain-counting method, an experimental emulsion, 'Type GO', has been produced. It differs from G5 emulsion in only one characteristic: its sensitivity.

GO emulsion is processed by exactly the same methods as G5, and it is therefore possible to use thick layers, and to make sandwiches of GO and G5 emulsion on the same glass base. As can be seen from fig. 1 (Plate XXVII*), the interface between the two emulsions is clearly defined.

The relation between grain density and energy loss in a batch of GO emulsion has been determined by grain counting in a sandwich plate consisting of 400 μ G5 emulsion covered with a 200-μ layer of GO emulsion. It is shown in fig. 2. The plate was developed with amidol-bisulphite (Herz 1952) to an average minimum grain density (in G5) of 26 grains/100 μ. The tracks used in the calibration were those of alpha-particles and singly-charged particles ending in the emulsion, taking all singly-charged particles to be protons.

Fig. 2



Grain density as function of energy loss in GO emulsion. The errors are standard deviations assuming a Poisson distribution. The grain densities to be expected for various relativistic nuclei are indicated.

GO emulsion has been designed in order to simplify the work of identifying heavy nuclei, to increase the accuracy for a given track length, and to increase the emulsion-to-glass ratio in the stacks of plates. It appears to satisfy all these requirements. Moreover, in addition to its application in cosmic-ray research, it may well be useful in nuclear physics, because it is an emulsion of low sensitivity which can be used in very thick layers, either alone, or sandwiched with layers of electron-sensitive emulsion. Work is continuing on materials of this type and it may be found desirable to produce emulsions of somewhat different characteristics.

---

*For Plate see end of issue.

References

Blau, M., Rudin, R., Lindenbaum, S., 1950, *Rev. Sci. Instrum.*, **21,** 978.
Bradt, H. L., and Peters, B., 1948, *Phys. Rev.*, **74,** 1828 ; 1950, *Ibid.*, **80,** 943.
Ceccarelli, M., and Zorn, G. T., 1952, *Phil. Mag.*, **43,** 356.
Herz, A. J., 1952, *J. Sci. Instrum.*, **29,** 60.
Hoang Tchang-Fong, 1950, *Thesis, Paris.*
Sørensen, S. O. C., 1949, *Phil. Mag.*, **40,** 947.

LV. *Notices of New Books and Periodicals received*

Linear Computation. By PAUL S. DWYER. (John Wylie & Sons Inc.)

THIS book deals with the solution of sets of linear algebraic equations and with related problems such as the calculation of the inverse of a matrix and the formulation of the characteristic equation. It is concerned almost exclusively with ‘ direct ’ methods, i.e., methods equivalent to eliminating the variables in turn, and iterative or other indirect methods are referred to only briefly. The book starts with a discussion of the approximate nature of the numbers used in most numerical calculations and the author defines five or six kinds of approximate number. While this discussion may be useful in clarifying ideas it cannot really be said to lead anywhere since once the author gets down to business he follows the general practice of using rounded-off numbers in which the error is unspecified and usually unknown. The book gives a minute account of the different methods with their variants which have been proposed for the solution of sets of linear equations. This involves much detail since differences in, for example, the order in which the multiplications or divisions are performed can make all the difference between a good and a bad numerical procedure.

The author’s own interest in the subject began in the field of statistics and he explains in the preface that the book is designed for readers with very little mathematical equipment. Determinants and matrices are not used in the earlier parts of the book but they are defined and their properties are developed *ab initio* in the latter part. The reviewer cannot help feeling that those who wish to go into the subject in the detail offered in this book could be expected to have more mathematical background and would have preferred a more concise treatment. Such a treatment would also have been more useful to people needing practical advice about how to solve a set of equations ; at present such people are likely to be bewildered by the amount of material in the book, and they will have to search to find the author’s conclusions about the relative advantages of the various methods he gives. The book is well provided with references and exercises for the reader to work. M. V. W.

Dielectric Breakdown of Solids. By S. WHITEHEAD. [Pp. 271.] (Oxford University Press.) Price 25*s*.

THIS is a most valuable book on dielectric breakdown. It contains a great amount of material of considerable interest to the engineer and the experimental physicist working in this field. There are six chapters respectively on method of treatment, intrinsic breakdown, thermal breakdown, breakdown caused by discharges, electro-chemical deterioration and on dielectric breakdown in practice. The reader interested in the fundamental side of this field will welcome the inclusion in Chapter II of a certain amount of theory, mainly based on Fröhlich’s work. Theory has accounted for a number of features of breakdown in a qualitative way and in some cases reasonable quantitative agreement with experimental data had been found. Nevertheless further, well controlled experiments, in particular on time effects of pre-stressing on intrinsic breakdown are wanted in order to obtain a more complete picture. The great difficulties in the necessary experimental technique as well as in the interpretation of the results will certainly be appreciated by readers of this monograph.

Finally, it should be mentioned that the material in the present volume on dielectric breakdown is almost completely different from the contents of the earlier (1932) book on the same subject by the same author.

D. P.

X-Ray Analysis of Crystals. By J. M. BIJVOET, N. H. KOLKMEIJER and C. H. MACGILLAVRY. (Butterworths Scientific Publications.) Price 50*s*.

THIS volume is the first English edition of a well known Dutch work published in 1938 and revised in 1948. For descriptive convenience only, its subject material may be most easily compared with ' The Crystalline State—Vol. I ' by W. L. Bragg but differs chiefly in respect of its simpler and less detailed approach, and by way of its greater emphasis on structural crystallography ; also the more recent material and references which it contains. As such it is a useful comprehensive text-book for the science student whose course includes crystallography, and also for the research beginner.

Starting with a simple introduction to x-ray diffraction directions, and crystal symmetry, it leads to the intensity of diffraction structure determination by trial and error and Fourier methods. In order to keep the text as lucid as possible, some of the details of the subject, such as the concept of reciprocal space, line-broadening etc., are presented in a fairly lengthy appendix.

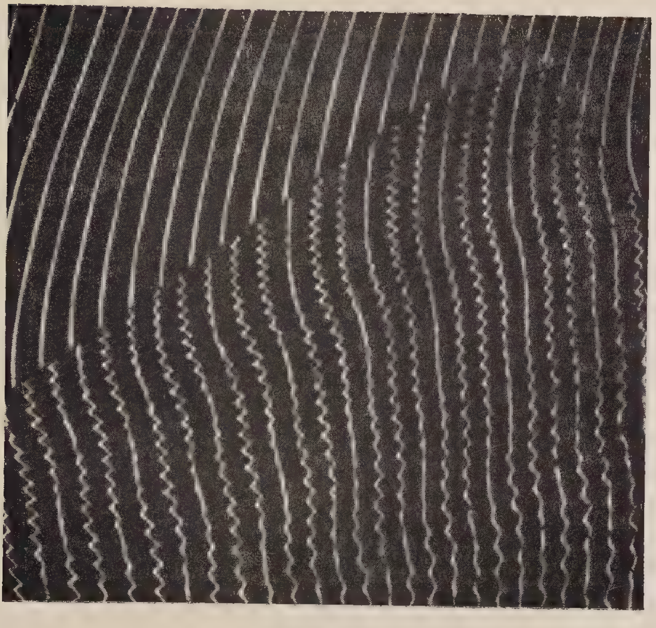
The second, slightly shorter portion of the book, is perhaps the most successful and is devoted to an excellently illustrated explanation of basic structure types. After a brief classification by bond-type, and a discussion of simple structure and the concept of atomic size. the dihalides. silicates and long chain compounds are emphasized as examples.

It is difficult to keep a comprehensive account of any subject up to date, and it is encouraging to meet reference to modern researches such as dislocation in metals, haemoglobin and neutron diffraction to quote just a few.

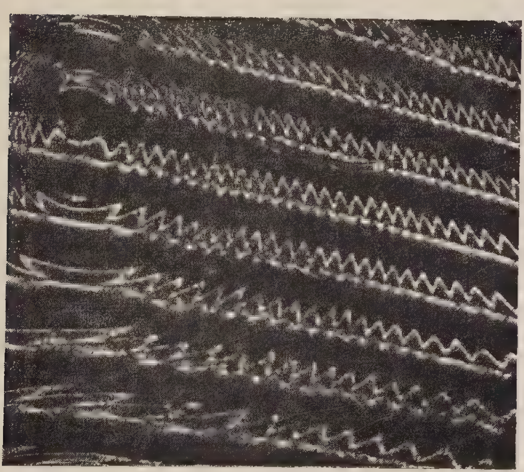
Although the text is reasonably complete in itself. the more serious worker is presented with twenty or so references at the end of each chapter, a fair proportion of which are later than 1945. This ample documentation makes the book interesting to the more serious student of the subject.

H. F. K.

[*The Editors do not hold themselves responsible for the views expressed by their correspondents.*]

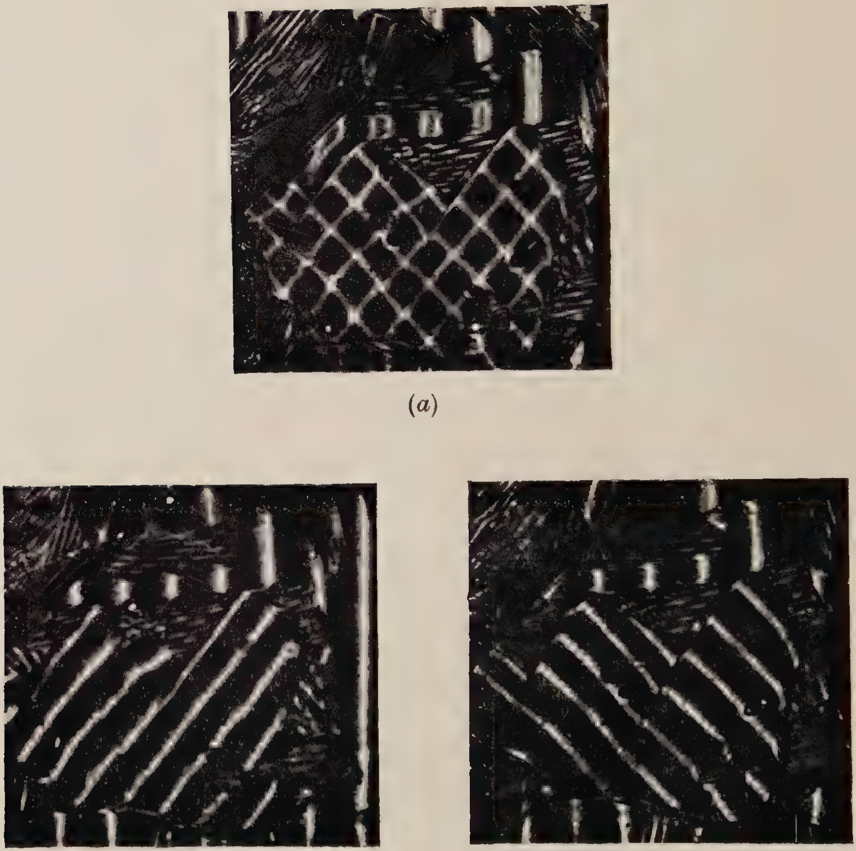


(*a*)



(*b*)

Fig. 1



(a)

(b)

(c)

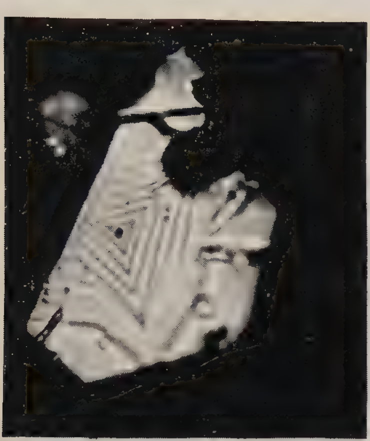


Fig. 1 (×1300)

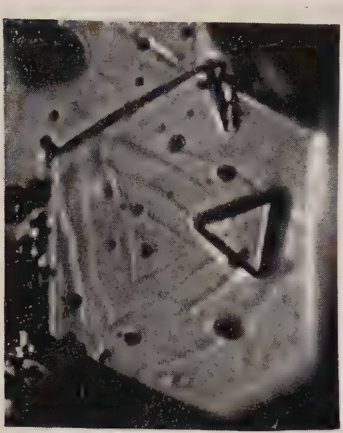


Fig. 2 (×1300)

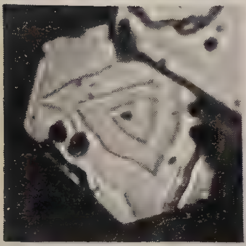


Fig. 3 (×1350)

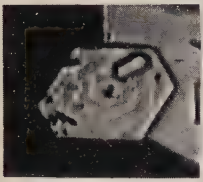


Fig. 7 (×1500)

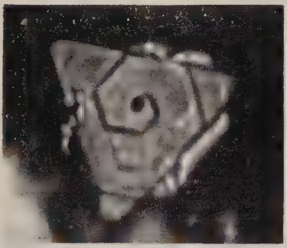


Fig. 4 (×1600)

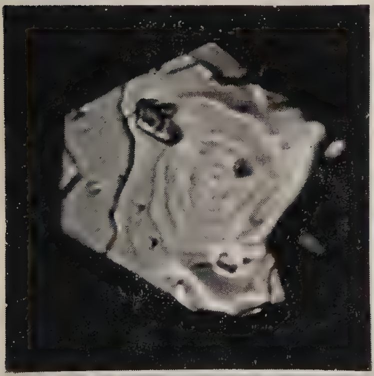


Fig. 5 (×1475)

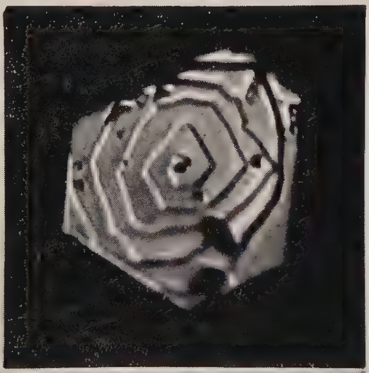
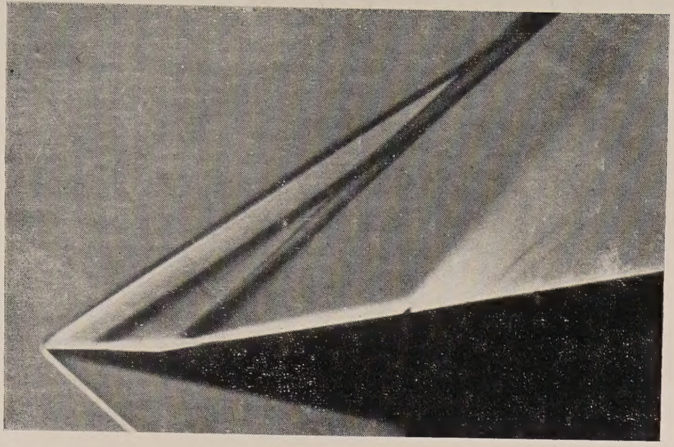


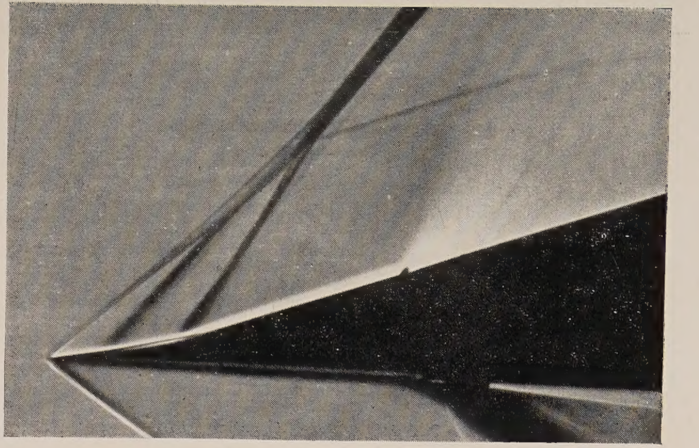
Fig. 6 (×1500)

Growth spirals on the (111) faces of gold crystals obtained by precipitation.

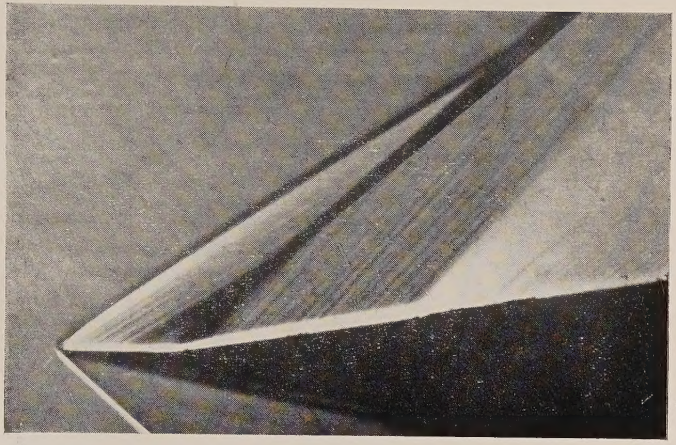
9



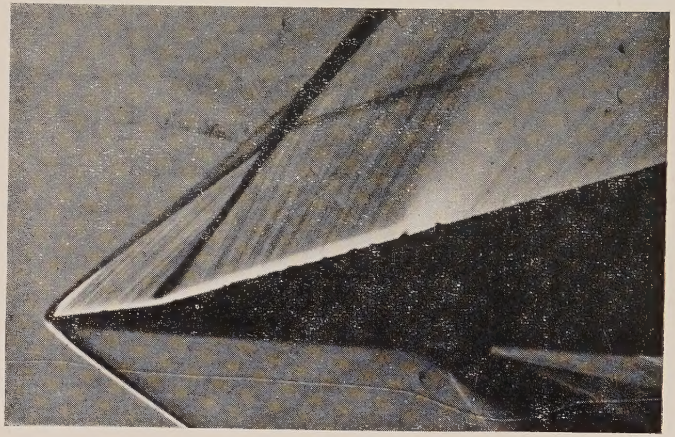
10



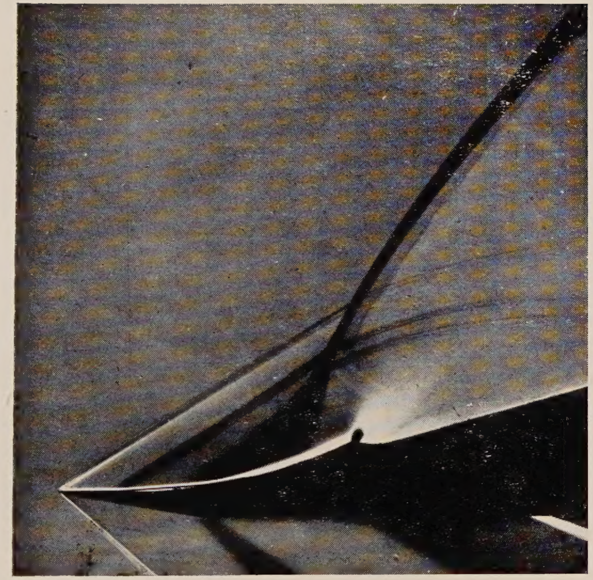
11



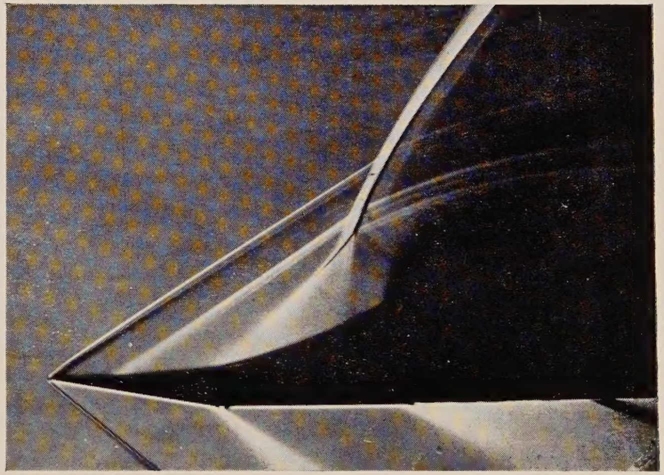
12



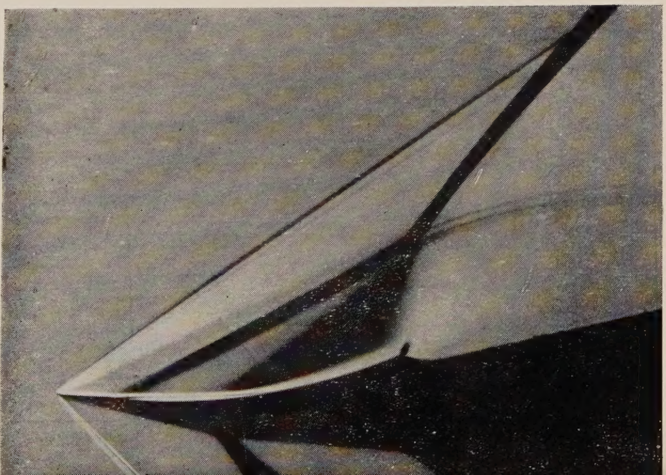
14



15



16



17

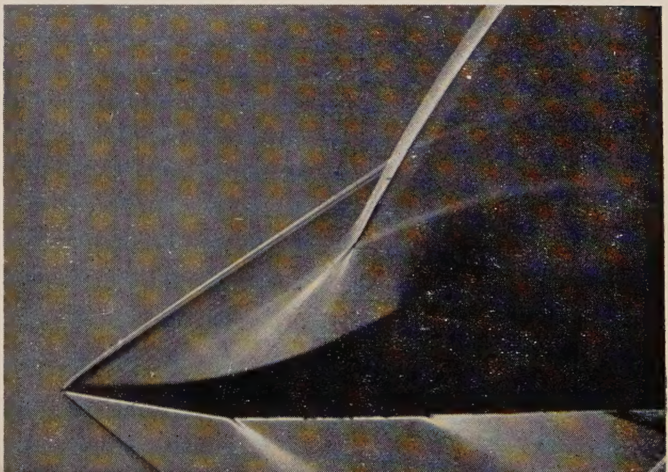
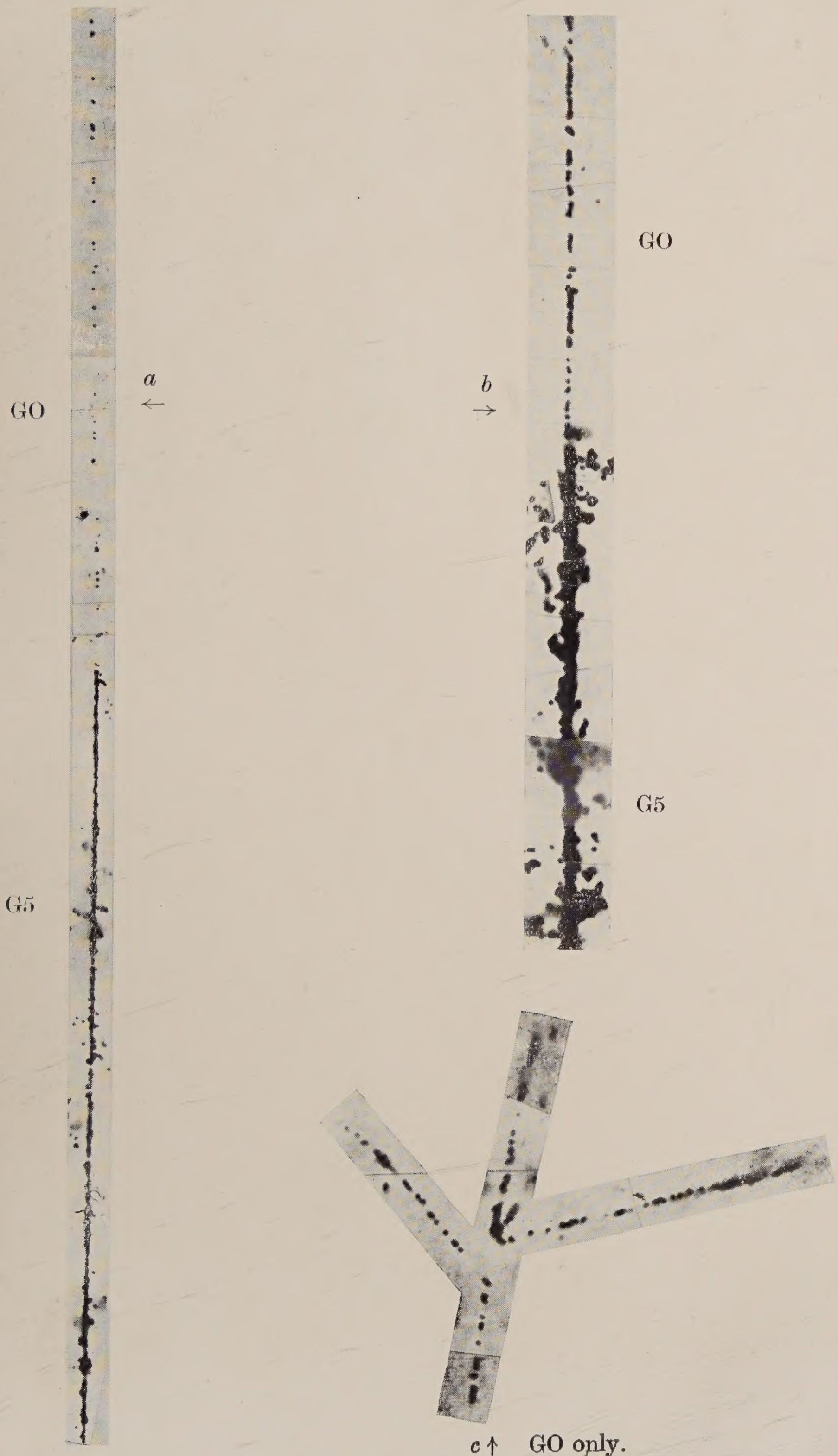


Fig. 1



Appearance of Tracks in G5/GO Sandwich Emulsion. *a* and *b* are heavy nuclei, *c* is a radio-active star. The scale of *b* and *c* is the same, and different from that of *a*.